



Machine Learning in Stem Cells Research

Zaman, W. S. W. K., Karman, S., Ramlan, E. I., Tukimin, S. N., & Ahmad, M. Y. (2021). Machine Learning in Stem Cells Research: Application for Biosafety and Bioefficacy Assessment. *IEEE Access*, 9, 25926-25945. [9344621]. <https://doi.org/10.1109/ACCESS.2021.3056553>, <https://doi.org/10.1109/ACCESS.2021.3056553>

[Link to publication record in Ulster University Research Portal](#)

Published in:
IEEE Access

Publication Status:
Published (in print/issue): 02/02/2021

DOI:
[10.1109/ACCESS.2021.3056553](https://doi.org/10.1109/ACCESS.2021.3056553)
[10.1109/ACCESS.2021.3056553](https://doi.org/10.1109/ACCESS.2021.3056553)

Document Version
Author Accepted version

General rights

Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact pure-support@ulster.ac.uk.

Machine Learning in Stem Cells Research: Application for Biosafety and Bioefficacy Assessment

Wan Safwani Wan Kamarul Zaman, Salmah Karman, Effirul Ikhwan Ramlan, Siti-Nurainie Tukimin, Mohd Yazed Ahmad

Abstract—The applicability of machine learning-based analysis in the field of biomedical field has been very beneficial in determining the biological mechanism and validation for a wide range of biological scenarios. This approach is also gaining momentum in various stem cells research activities, specifically for stem cells characterization and differentiation pattern. The adoption of similar computational approaches to study and assess biosafety and bioefficacy risks of stem cells for clinical application is the next progression. In particular where tumorigenicity has been one of the major concerns in stem cells therapy. There are many factors influencing tumorigenicity in stem cells which may be difficult to capture under conventional laboratory settings. In addition, given the possible multifactorial etiology of tumorigenicity, defining a one-size-fits-all strategy to test such risk in stem cells might not be feasible and may compromise stem cells safety and effectiveness in therapy. Given the increase in biological datasets (which is no longer limited to genomic data) and the advancement of health informatics powered by state-of-the-art machine learning algorithms, there exists a potential for practical application in biosafety and bioefficacy of stem cells therapy. Here, we identified relevant machine learning approaches and suggested protocols intended for stem cells research focusing on the possibility of its usage for stem cells biosafety and bioefficacy assessment. Ultimately, generating models that may assist healthcare professionals to make a better-informed decision in stem cell therapy.

I. INTRODUCTION

Stem cells are undifferentiated cells found in all multicellular organisms which possess a unique self-renewal ability and multi-potential differentiation [1-3]. Stem cells have been associated with the fields of regenerative medicine and tissue engineering with the goal to improve health and quality of life, especially patients with debilitating diseases. Stem cells can be divided into three categories: (1) embryonic stem cells (ESC) derived (ESC) from early-stage embryos; (2) adult stem cells (ASC) and (3) induced pluripotent stem cells (iPSC). These cells owe its regenerative capacity to its ability to migrate to the

injured part of the body, to divide and produce daughter cells, which have the ability to differentiate into other lineages of cells in order to repair the damaged tissue under appropriate conditions [4]. The ability for stem cells to induce regeneration can be influenced by culture condition and the type of secretomes released [5]. Stem cells have been studied and even applied for the treatment of various clinical conditions. However, there are risks which needed for further evaluation (before clinical application), such as miss-differentiation of cells, miss- targeting of cells, immune rejection and the biggest concern is genomic instability or tumor formation [6, 7].

While the application of stem cells for treatment is on the rise, their overall quantity in the body is scarce. Generally, cell therapy protocols require hundreds of millions of MSC per treatment and this would require cell expansion in vitro for about 10 weeks before implantation [8]. In this regard, long-term expansion or manipulation of stem cells may contribute to cellular senescence or even tumorigenesis in vitro, which may cause them to be non-viable for clinical usage. This has led to concerns of biosafety and bioefficacy of stem cells in clinical application [9, 10]. The aforementioned concerns are mainly due to poor understanding of stem cells biological mechanism, which has prevented it from being used widely in research in clinical application. Experimental approaches based on phenotypic and genotypic profiling are limited, whereby, they can also be expensive and time- consuming [11-13]. Furthermore, these approaches would also require subsequent validation assays to confirm its accuracy, which apart from the small sample size, can also lead to misinterpretation of data.

The recent development in stem cells research has shown that machine learning application can be used to overcome some of these limitations, particularly in phenotypic profiling of stem cells [14, 15]. Other potential applications that could be explored are annotation of stem cell genome [16], predictions of protein binding, identification of specific markers [17] or key transcriptional factors of stem cells and characterization of stem cells transcriptional regulatory networks [18].

There are datasets generated from experiments to quantify molecular variables related to stem cells biosafety and bioefficacy, such as the gene and proteins interactions. However, these datasets are complex with an intricate network of molecular interactions and analysis [19, 20]. To address this complexity, machine learning could provide next-level analyses

Wan Safwani Wan Kamarul Zaman, Mohd Yazed Ahmad, Salmah Karman and Siti-nurainie Tukimin are with Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

Wan Safwani Wan Kamarul Zaman, Mohd Yazed Ahmad are with Centre for Innovation in Medical Engineering (CIME), Faculty of Engineering, University of Malaya, 50603, Kuala Lumpur, Malaysia

Effirul Ikhwan Ramlan is with School of Computing, Engineering and Intelligent Systems, Ulster University, Magee Campus, School Office MS013, Northland Road, Derry/Londonderry, BT48 7L, United Kingdom.

Corresponding author: Salmah Bt Karman (salmah_karman@um.edu.my)

that would allow better insights and the generation of new information for better biosafety and bioefficacy assessment [18]. As such, this would also allow medical practitioners to be better informed in offering personalized treatment to patients in stem cells therapy. A similar approach is seen in cancer research, whereby, machine learning has been widely applied in the identification and classification of cancer cells. Similar machine learning models and approaches can be applied in stem cells research [21-24] that could assist in accelerating the evaluation of safety and efficacy of stem cells. This could potentially bring stem cells to the forefront of personalized medicine. The current and future research trends in stem cells research are presented as an overview in Figure 1. In this review, the advantages and limitations of machine learning in stem cells research were presented, including on how next-generation machine learning methods could be used to expand our understanding of stem cells biology and their biosafety and bioefficacy risks. We anticipate that machine learning could have substantial impacts on stem cells research and therapy, providing a supporting tool in making a personalized clinical decision. This includes tailoring treatments for optimization in individual patients.

II. STEM CELLS BIOSAFETY AND BIOEFFICACY PROFILES

In general, stem cells are unspecialized cells with self-renewal and differentiation (into specialized cell) abilities [25]. However, each type of stem cells has different characteristics, which are attributed by their origin, biological characteristics and functionality. The ESC cells are pluripotent stem cells originated from the inner mass of blastocyst of the embryo and can give rise to the entire body tissue organs except for placenta and umbilical cord [26-28]. Meanwhile, ASC cells are somatic cells-derived from a certain part of the adult body, which can only give rise to stem cell progeny of the original site [29] in which they are found. They are known to be multi-potent with limited differentiation capacity [30, 31]. Adult stem cells are also known as mesenchymal stem cells (MSCs) [32-35].

[Fig. 1 about here.]

The iPSCs are capable in giving rise to all kinds of cell types in the body but the difference is that iPSCs are reprogrammed stem cells, whereby, somatic or primary cells are biologically reprogrammed, giving rise to stem cells similar in characteristics as ESCs in culture [36-38]. Due to this technique, iPSCs have been widely used as it reduced the dependency on ESCs and ASCs, which are limited in the cell population. The iPSCs have been profoundly utilized not only for repair, replenishment and replace the damaged cell, tissue and organ but they have also been employed for drug-response therapy [39]. Due to the regenerative capacity of stem cells, they have been regarded as a powerful tool in regenerative medicine, particularly in the treatment of debilitating diseases. However, such potential and capability have given rise to other clinical concerns, such as adverse effects associated with

biosafety and bioefficacy issues. These effects may not materialize immediately after receiving stem cell therapy. Post therapy monitoring may be difficult as there are no established pre- or post-parameters and further, there is no 'one-size-fits-all' protocol to enable such monitoring procedure. To develop a stem cell-based therapy, we must first ensure the safety and efficacy of stem cells. High efficiency of stem cells is needed to have effective homing, engraftment and persistence in damaged tissues, which would enable a stable interaction between the transplanted and the injured tissues. This is important to maximize the therapeutic capacity of stem cells.

Studies on biosafety and bioefficacy of stem cells have been carried out for many years. There have been few reports addressing the biosafety and bioefficacy profiles of stem cells [10, 40, 41], which showed the importance of addressing these issues. As of now, there is no standard or conclusive data that can be used to establish a proper protocol for biosafety and bioefficacy assessment. Further, any protocols and guidelines established would need to be internationally accepted and harmonized [42]. The proposed minimal criteria to define human mesenchymal stem cells (MSC) was established by the Mesenchymal and Tissue Stem Cell Committee of the International Society for Cellular Therapy (ISCT). These criteria are; 1) MSC must be plastic-adherent when cultured in standard culture conditions, 2) MSC must express CD105, CD73 and CD90 and lack in the expressions of CD45, CD34, CD14 or CD11b, CD79 α or CD19 and HLA-DR surface molecules and 3) MSC must be able to differentiate to osteoblasts, adipocytes and chondroblasts in vitro [43].

Since then, further investigations on stem cells characteristics were carried out, which can be (directly or indirectly) used to evaluate the biosafety and bioefficacy profiles of stem cells. With regards to adult stem cells (ASC) or MSC, studies have shown that these stem cells were reported to have low risks of tumorigenicity in long-term culture [4, 44] and low risks of abnormalities following long-term cryopreservation [45]. While there are no significant changes in stem cells differentiation ability, cryopreservation caused stem cells to appear less fibroblastic in appearance [46]. Long-term culture of stem cells was also reported to alter its stemness and differentiation ability [46, 47]. On the other hand, tumorigenicity risks of embryonic stem cells (ESC) and induced pluripotent stem cells (iPSC) have been reported, which pose a hurdle in stem cell therapy [48, 49]. Furthermore, the differences in cell microenvironment and culture conditions contributed by biophysical and biochemical cues can affect stem cells response, for example, cell culture in hypoxia [50-53], use of serum [54] and fluid shear forces [55]. This evidence also showed that stem cells response is affected by both static and dynamic interventions.

Despite the successful clinical application of stem cells, the sample size is rather small, which may not be sufficient to determine the safety and efficacy of the treatment. The application of human allogeneic adipose-derived MSC showed feasibility in a pediatric patient with no adverse effects of up to 12 months following treatment [56]. While Lennmyr et al. [57] showed that all adult patients affected by lymphoblastic

leukemia was successfully treated with allogeneic hematopoietic stem cell transplantation (alloHSCT) with an increased overall survival rate of more than 10% in 5 years [57]. Meanwhile, Schlenk et al. [58] evaluated alloHSCT among patients having acute myeloid leukemia (AML) showed a significant beneficial impact after treatment. Similarly, Cornelissen et al. [59] reported that AML patients that were treated using alloHSCT have a significant beneficial impact with the overall increased survival rate of 12% and was successfully validated using cytogenetic profiling. However, there have been variations in terms of treatment outcomes and responses, which can be due to multiple factors that can be difficult to ascertain. Examples of these factors include different donors [60], age of donors [61] and different culture protocol [62], which may contribute to non-standardized outcomes and potentially adverse effects in stem cells therapy that can be irreversible. This has made it difficult to assess and to establish a standardized protocol to evaluate the biosafety and bioefficacy risks of stem cells. Therefore, machine learning-based predictive analytical methods are desirable to accelerate the discovery of new stem cell markers for safety assessment and to forecast stem cell therapy efficacy in order to minimize the potential adverse effects and to maximize the success of treatment.

III. LIMITATIONS IN STEM CELL THERAPY WHICH POSE EFFICACY AND SAFETY ISSUES

Although suitable stem cells safety and efficacy profiles and assessment are still not well established, stem cells have been used for various disease treatments. Stem cell therapies have been applied for the treatment of anemia [63], multiple myeloma [64-66], arthritis [67, 68] and even stroke [69, 70]. Stem cells therapies have also been applied for blood-related cancers, whereby, patients have undergone allogeneic hematopoietic stem cell transplantation (alloHSCT) for thalassemia [71] and acute myeloid leukemia (AML) [72-75]. In some cases, determining post-treatment efficacy and safety of stem cell therapy may be restricted due to the difficulty in following up after treatment. It might also be due to stem cells' dynamic responses in different individual recipient and disease models, which lead to variations in the outcome of the treatment. In this regard, there have been reports addressing post-treatment complications of stem cells therapy within a year following the treatment.

Rovo and Tichelli [76] reported cardiovascular complication risk following allogeneic hematopoietic stem cell transplantation leading to considerable morbidity and mortality, including patients having critical diseases, such as dyslipidemia, arterial hypertension, diabetes mellitus and kidney disease. There have been reports of undesired differentiation and malignant transformation [77] as well as the ability to promote tumor growth and metastases, which has been a major concern in stem cells therapy [78]. Patients who have undergone autologous stem cell transplantation (ASCT) for lymphoma have a significant risk of developing therapy-related acute myeloid leukemia [74, 79, 80] and

myelodysplasia (t-AML/MDS) [81]. This may be attributed by the ASCT procedure that includes priming chemotherapy, total body irradiation and the extensive cellular proliferative, which occur during engraftment, leading to the development of t-AML/MDS. Graft-Versus-Host Disease (GVHD) is another adverse effect that occurs following stem cell treatment [82-84], which can be overcome by the use of a mismatched allograft that necessitates T cell depletion. Surprisingly, a greater HLA mismatch was associated with a lower risk of GVHD [85]. The mismatched donor lymphocyte infusion (DLI) was specifically created for prophylactic treatment of T cell depleted mismatched allograft recipient [85].

All of these stem cells donors and recipients' responses may be important factors needed to be considered carefully. Biological assays and genetic molecular expression data profiling may be able to overcome such limitations and challenges, but they may be cost-prohibitive and time-consuming. Antibiotic matching, biomarker details and signaling pathways are all essential information needed but may require subsequent validation assays for accuracy. In this regard, dependency on biological assays may lead to the misinterpretation of data, particularly in terms of the similarity in biomarkers and molecular signaling pathways of various microenvironment and disease models. Hence, machine learning-based characterization and classification profiling techniques may be able to capture the genotypic and phenotypic differences as well as the changes that occur in a shorter period with more accuracy in terms of safety and efficacy of the stem cells.

IV. MACHINE LEARNING APPLICATION OPPORTUNITY IN STEM CELLS BIOSAFETY AND BIOEFFICACY EVALUATION

Understanding stem cells behavioral response and changes have been mainly carried out through biological assays that employed time-consuming and laborious methods [86, 87]. Furthermore, just like any biomedical datasets, stem cells datasets are generally limited by sample size [88]. To increase data for analysis, some investigators used 3D printing to create bio-scaffolds to mimic the natural environment of stem cells, but this approach was not always successful, whereby, stem cells often migrated away from the printed scaffolds or locations. Hence, the analysis of stem cells' intrinsic ability and response were not always accurate. To overcome this limitation, the machine learning approach to study stem cells' complexity is gaining momentum, particularly on the aspect of molecular and genomic changes in pluripotent stem cells.

Machine learning is a common method in data analytics for identification and recognition of patterns, which when applied to stem cell biology, will enable the discoveries of new insights with reasonable accuracy in a shorter amount of time. This approach would also be of advantage as stem cells are known to form predictable patterns in their natural environment as they mature into tissues. Such analytics are difficult to duplicate in the lab, costly to perform, laborious and time consuming to execute. For instance, Libby et al. [89] used extended cellular Potts model to capture pluripotent stem cell organization

dynamics that enabled them to demonstrate morphogenic dynamics through a model-driven exploration of stem cells behaviors, which is a vital step in organ modelling [89]. Understanding that the individual stem cell is different even if they are genetically cloned has led to the Allen Cell Explorer produced by the Allen Institute for Cell Science in Seattle, USA. The Allen Cell Explorer, which complements various ongoing projects, is an online catalogue, including the 3D images of stem cells as well as the iPSC that were produced using deep learning analysis and cell lines altered with the gene-editing tool, CRISPR [90]. It is also a growing library that charts the uniqueness of single cells at DNA, RNA and protein levels [90]. This gives a more holistic and unbiased approach to predict and understand multiple aspects of cellular structure and behaviors.

Although machine learning application in stem cells research is not a commonality at this stage, a proof-of-concept study has been presented previously [91]. Zhang et al. [91] employed machine learning and microscopic image analysis to identify iPS progenitor cells in their effort to understand the origin and underlying mechanism of iPSC particularly at the early stage of cell reprogramming, including the biomarkers involved. However, the proposed model by Zhang et al. [91] showed inconsistencies in their prediction with large fluctuation. The model can only predict iPS progenitor cells with a minimum precision of 52%. The model is incapable of handling additional iPSC features and phases to achieve higher accuracy of the prediction performance.

Machine learning models have been applied in cancer diagnostics and prognostics [15, 92-94], whereby, similar predictions and interpretation models can be applied in understanding stem cells, specifically for its biosafety and bioefficacy evaluation. Similar machine learning techniques can be used in predicting and studying the dynamic changes of stem cells behavior in a particular environment, which should be directed towards understanding its impact on biosafety and bioefficacy for clinical application. The combination of the dataset from cancer research and stem cells research for various machine learning models should be considered as an approach in understanding stem cells behavior and interactions as well as the risks of developing adverse effects following therapy.

1. Image-based Dataset for Machine Learning in Stem Cells Research

The common workflow of image processing consists of the preparation of image input data, pre-processing, segmentation, feature extraction and classification steps. The microscopic image should be prepared by capturing the cell image from cells or tissue samples using the microscopic digital camera or software. Technical expertise for manual classification may be required for image labelling, especially for images that will be used for supervised training. For automated application, the microscopic images provided by authorized databases can be used for training and testing. For simple and efficient methods of segmentation and feature extraction steps, techniques such as Convolutional Neural Networks (CNN), K-Means and Mean

Shift should be adopted. The SVM, Naive Bayes and Fully Convolutional Neural Network (FCNN) are among the techniques that have been used for cancer cell classification, which has the potential to be applied in stem cell research. Image processing or analysis has been beneficial in cancer research. Computer vision software based on machine learning and deep learning algorithms is making automated analysis possible in delivering fast and accurate results. In this regard, image processing plays a crucial role in the diagnosis and detection of cancers as well as in monitoring cancer progression patients [23].

Similarly, machine learning is capable in overcoming limitations in stem cells research, whereby, lab-based characterization and classification using chemical reagents and biological assays can be labor-intensive, expensive, and time-consuming as well as less accurate. Most implementations of supervised machine learning relied on extensive training data using extracts from large and high-throughput biological data and features, such as cellular images and genome analysis. From the perspectives of cancer diagnosis and stem cells therapy, the machine learning approach is useful to understand the regulatory genomics. This includes the identification of regulatory variants, the effects and origins of mutation using DNA sequence, analyzing whole cells, the population of cells and tissues through detecting features that can be difficult or impossible to uncover in conventional laboratory settings [95]. Pattern recognitions and classifications of such biological data are important in identifying factors, which pose biosafety and bioefficacy risks of stem cells in clinical application.

While different techniques have been developed for analysis, deep learning method provide a more effective strategy due to the diversity of the data. It has been used to classify lesions and nodules; localize organs, regions, landmarks and lesions; segment organs, organ substructures and lesions; by retrieving images based on content; generating and enhancing images; and combining images with clinical reports [13, 96, 97]. The application of deep learning in analyzing images has been widely used in cancer stem cell phenotype research. Ke Fan et al. [98] demonstrated that the combination of SVM, RF and CNN was able to measure morphological dynamic and colony formation of iPSCs within 7 days. The application of SVM by Tanaka et al. [99] enabled automated classification of adipogenic and osteogenic differentiation as well as undifferentiated features of human mesenchymal stem cells (hMSC) in RGB color image. Meanwhile, Theagarajan and Bir Bhanu [100] have developed and proposed new automated detection and classification of human embryonic stem cell (hESCs) with an accuracy of 94.46% using the application of CNN for phase contrast hESC image analysis.

2. Stem Cells and Cancer Cells Feature Engineering for Machine Learning

Several studies have shown that stem cells and cancer cells share some similarities. These similarities can be attributed to their functional capability that is conceptually similar in terms of their ability to self-renew and to proliferate [101]. However,

they are also fundamentally different, whereby, these cells can be distinguished by different regulatory mechanisms reflected in at least three characteristics; 1) propagation and proliferation ability, 2) morphology and 3) cell surface markers. These characteristics may be considered for risks evaluation associated with the safety and efficacy of stem cells. Image-based high-content screening has also become increasingly important in stem cells research in monitoring the changes in phenotype, such as cell morphology and differentiation [102, 103].

2.1. Proliferation and propagation ability

In terms of proliferation ability, it is important to understand that normal stem cells are regularly more vigilance in controlling their proliferation, but this ability is lacking in cancer cells [101]. Cells population doubling time is one of the distinguishing factors that can be included as a feature used to train the machine learning model. However, it is important to also take into account the different models [101] used to understand cancer propagation. Cancer stem cells (CSCs) model is rare and is a phenotypically distinct group of cells, which may hierarchically induce the stable generation of non-tumorigenic and tumorigenic cells. They can probably be generated from normal stem cells or precursor cells within tissues after mutations and resistant to conventional chemotherapy occurred [104].

Although they are rare, certain markers have been suggested for the identification of CSCs. In the clonal evolution model, cancer cells are distinctive in phenotypes with malignant potential and ability for disease propagation by undergoing additional genetic mutations. While in the interconversion model, cells can interconvert between being actively malignant and relatively quiescent, which is associated with the phenotypic differences between these cells. Although there are distinctive differences in the cells in each model, they are not mutually exclusive, whereby, tumorigenic cells are able to undergo further genetic and epigenetic alterations, depending on their microenvironment regardless of which model the cells follow [105]. Based on the propagation and proliferation ability, specifically associated biomarkers, cell count or numbers and time-period depicting cells population growth could be the features that can be included in the training models to distinguish the characteristics of normal stem cells and cancer-associated cells.

2.2 Cell Morphology

Morphologically, stem cells and cancer cells may show different cell features, which can be viewed microscopically (Figure 2). Microscopic images of cells are crucial to extract information for the machine learning models to distinguish the different features of normal stem cells and cancer cells based on their sizes and shapes. Generally, the appearance of normal stem cells is more consistent in their shapes and smaller in size while cancer cells can be abnormal and vary in shapes and sizes, which may be contributed by their heterogeneous nature. Although they may exhibit unique differences in their morphology, cell features, and motions require laboratory

experimental approaches to create sample dataset prior to machine learning modeling. Based on a suitable model, the identification of cancer progenitor cells can be confirmed based on the morphology and motion pattern, which may be different from normal cells.

Zhang et al. [91] used time-lapse microscopic images of iPS forming cells in early stage reprogramming and selected 11 types of cell morphological and motion features, which included the area of coverage and speed for modelling to perform feature selection. Further analysis of cell motion showed that migratory motions for progenitor cells can be distinguished by the direction and distance to bring distant progenitor cells together. However, the input of cell features described by Zhang et al. [91] and Meygola et al. [106] would require high-resolution time-lapse imaging to allow the detection or tracing of cellular events. With regards to cell segmentation and tracking, Dzyubachyk et al. [107] used coupled active surfaces algorithm and time-lapse fluorescence microscopy images. While Türetken et al. [108] proposed an integer programming to track elliptical cell populations in time-lapse image sequences. In the case of image segmentation, the challenge with live-cell imaging is in determining which parts of images correspond to which individual cells. Van Valen et al. [83] showed that this can be solved by applying CNN that can robustly segment fluorescent images of cell nuclei and phase contrast images of cells without the use of a fluorescent cytoplasmic marker.

[Fig. 2 about here.]

2.3. Cell surface markers

Cell surface markers are associated with features and changes in cell morphology and progression. Some require deep epigenetic experimental approaches for input. It is challenging in determining specific markers for normal stem cells, cancer cells or CSCs as most of these markers can be presented in all types of cells, making them non-specific. Otherwise, these aspects would require machine learning approach in identifying specific cell surface markers. In comparing between normal stem cells and cancer cell progression, it is mostly discussed within the context of CSCs, as stem cells have also been shown to be involved not just in cancer initiation and progression but also in CSCs generation [109]. Nevertheless, contradictory results on CSCs and stem cells relationships are still very much debatable.

Although both cells share some similarities in terms of self-renewal and differentiation ability, there are studies. Some studies showed different characteristics between the two cell types and this can be delineated by the existence of specific cell markers [101, 110]. This feature can be used for machine learning approach to classify and track stem cells progression in a particular environment for the risks of them conforming or inclining towards cancer-associated cells. An important attribute of CSCs is that they have the ability to trans-differentiate into different phenotypes [111, 112], whereby, they can express angiogenic and vasculogenic markers and also

be able to organize a pseudo vascular network. Several studies have also associated these characteristics to the expression of potentially CSC markers in several types of cancer cell lines, such as breast cancer cell lines, MDA-MB 453 and MDA-MB 231 [113-115], non-small cell lung cancers (NSCLCs) [116], renal cell carcinoma Cell [117], nasopharyngeal carcinoma cell (NPC) [118], colon cancer cells mucopidermoid carcinoma cell lines (YD15) and its derivative (YD15M) [119]. From these studies, it can be summarized that the characteristics of malignancy and cancer progression were typically associated with a panel of surface markers, which are $CD133^+$, $CD44^+$, $CD24^-$, $OCT3/4$, or/and $NANOG$ [120, 121]. On the other hand, the $CD44^{high}ESA^{lower}$ or $CD44^{high}ESA^{high}$ expressions indicated the presence of CSCs population in squamous cell carcinoma in breast through a comprehensive analysis of data obtained from flow cytometry, immunohistochemical and real-time polymerase chain reaction (RT-PCR) [112].

Surface markers regulation leading to the induction of epithelial-mesenchymal transition (EMT), which resulted in the acquisition of invasive and metastatic properties is also one of the characteristics found in CSCs [119, 122]. EMT phenomenon in CSCs has been reported as metastasis precursor, which enable the cells to acquire invasiveness and become extremely resistant to conventional therapies [112, 117]. The down-regulation of E-cadherin and upregulation of N-cadherin, which are termed as cadherin switching cascade, is a major hallmark of EMT. The cells which are undergoing EMT can be accurately identified through intensive genomic profiling for downregulation of cytokeratin (CK) and upregulation of vimentin, N-cadherin and fibronectin. This may be important markers for the characterization and identification of CSCs [119].

Currently, image analysis has been employed in the study of stem cells reprogramming and its progression using iPSCs. Kusumoto et al. [123] employed CNN to identify endothelial cells derived from iPSCs, whereby, the networks were trained using phase contrast images of endothelial cells based on morphology only. The network performance was then assessed by K-fold cross-validation, which confirmed that CNN was able to identify endothelial cells based on morphology with high performance. On the other hand, computer vision-based deep learning was also used to study the progress of stem cells differentiation. CNN was also able to be trained with transmitted light microscopy images to identify pluripotent stem cells from early differentiating cells and its ability to recognize the features with more than 99% accuracy [124]. Similarly, the classification of light microscopic images was used to predict lineage choice and cellular movement of primary hematopoietic progenitors during differentiation [125].

Despite limited machine learning application for stem cell biosafety and bioefficacy, comparative and classification analysis of stem cells can be carried out by comparing the cell images of stem cells and cancer cells without having the dependency on molecular and biological assays. Depending on a particular niche (i.e., whether in vitro or in vivo conditions), stem cells can initiate or acquire senescence or cancer

characteristics. This has been demonstrated in glioblastoma multiforme study by Adamski et al. [126], which reported that, there is a putative link between cellular dormancy of malignancies and stem cell-like characteristics in cancer that could be due to the co-expression stem cells markers. Based on these studies, the prediction in the risks of stem cells to acquire cancer characteristics prior to clinical applications is possible.

V. TECHNICAL RECOMMENDATION OF MACHINE LEARNING IN STEM CELLS RESEARCH

Machine learning classification techniques have been applied in cancer research to identify and classify the types of cancer cells with relatively high accuracy, sensitivity and specificity. Some popular applications involved Support Vector Machine (SVM), K-Nearest Neighbors (KNNs), Artificial Neural Networks (ANNs), Decision Tree (DT), Random Forest (RF) and Bayesian Networks (BNs) [121, 122, 127-131]. In addition to cancer, the classification of microscopic red blood cells images from hematological disorder, such as sickle cell disease using deep-CNNs were able to reveal a diverse and any alteration in the cell shapes related to their biomechanical and bio-rheological characteristics. The deep-CNN employed showed good performance, high accuracy and robust predictions that enabled clinicians to assess the severity of the disease [132]. Similar techniques can be used to assess or profile stem cells biosafety and bioefficacy based on image analysis.

[Fig. 3 about here.]

With regards to image analysis, there are typical preliminary steps of image processing techniques. The process consists of input image data, image pre-processing, image segmentation, feature extraction and classification steps (Figure 3). The microscopic data is commonly used as an input for the training. Segmentation is a critical step that ensures the success of all subsequent algorithm steps. This entire approach of segmentation can be divided into three steps: image pre-processing, image segmentation and features extraction. In image pre-processing, the process commonly starts with the selection of sub-image or panel (usually nuclei or cell). The RGB color components and grayscale analysis will then be performed for sub-image or panel to find the contrast among the sub-image objects. The contrast level of RGB components is shown by the colored histogram.

The further segmentation process will then be performed to the sub-images by doing image refining, filling and splitting processes to identify the region of interest of the respective sub-images. Features such as size, shape and texture can then be extracted according to the calculation on the region of interest. By comparison, between Mean Shifts segmentation [131] and K-Means segmentation [132], the K-Means technique is one of the most popular method as it is simple, fast and efficient. Technically, K-means makes two broad assumptions, whereby, the number of clusters is already known. K-means is fast and has a time complexity $O(knT)$, where k is the number of

clusters, n is the number of points and T is the number of iterations [133]. The results of segmentation can then be used for the classification process to differentiate between the normal and abnormal cell. The typical process of image processing shown in Figure 3 has been commonly used in most nuclei-based cancer image processing. For example, by applying deep learning, patterns from several types of data, such as from cancer cell dataset and stem cells dataset can be automatically extracted [133]. This includes the detection, segmentation and recognition of cell images that can be used to predict the risk of cell irregularities that could jeopardize stem cells clinical application. The detection of cell irregularities is a multistage process, which also includes pre-processing tasks, such as segmentation and feature extraction from microscopic cell images before the application of CNN [134]. This approach can be used to observe and evaluate stem cells progression, particularly during the expansion phase to detect and predict the risk of abnormalities prior to clinical application. The expansion of cells is required to increase the number of cells to ensure sufficient cells can be used in stem cells therapy. It is a cell manipulation procedure, whereby, technical manipulation can increase the possibility of genotypic and phenotypic alterations [135].

Naik & Dixit [134] reported detailed technical steps in detecting cancer from microscopic biopsy images, comprising of the training and testing of the algorithm model. The machine learning architecture of cancer detection by Naik & Dixit [134] is shown in Figure 4. For both training and testing tasks, the sequence of step started by taking image samples using a microscope, followed by the segmentation and features extraction step using the CNN based image processing and finalized by a classification step using the Naive Bayes Algorithm. By using the CNN based image processing, the microscopic image that contains nuclei, cytoplasm and other features are segmented into 12 smaller bricks. In each segmented brick, the CNN based interpretation on types of cancer was done according to the features of grey level, color, texture, Law's Texture Energy (LTE), wavelet and Tamura's features. This interpretation, which was given in percentage, was then subjected to Naive Bayes algorithm to classify whether the image indicates the cells to be cancerous or not. In this regard, the CNN algorithm may be applied as the basic principle of deep learning-based cell identification. As reported by Kusumoto et al. [123], the deep learning identification is more straightforward and achieves higher accuracies compared to the other machine learning techniques without the requirement of image labelling. The technical steps implemented by Naik & Dixit [134] can be adopted for the detection and evaluation of stem cells biosafety and bioefficacy risks.

[Fig. 4 about here.]

In cancer research, the focus on early detection is important to stop or slow down the progression of tumor growth. Similar

motivation can be applied in stem cells research, whereby, machine learning approach can be used to observe stem cell progression, especially during the expansion phase to evaluate the risk of biosafety and bioefficacy. To do this, there should be an automated identification system based on cell morphological images using the machine learning approach. With the enhancement technology of CNN and deep learning algorithm on cell image analysis, the nuclei-based analysis seems promising. The nuclei-based analysis has been used in cervical cancer and blood-related cancer screening. The images can be obtained or captured using a light microscope supported by a CCD camera in a standard size. These images can be included as a dataset, which can technically be classified into categories for labeling (such as normal, abnormal, healthy or unhealthy).

These datasets can be applied for automated machine learning approach for abnormality evaluation of stem cells. Hussain et al. [136] has simultaneously conducted nuclei (nucleus) segmentation and classification from the cervical cancer morphological cell image using U-net architecture-based fully convolutional neural network (FCN). Figure 5 shows the nuclei image processing and machine learning classification architecture by Hussain et al. [136]. They adopted the shape representation model based on auto-encoders which act as a network regularizer to increase the strength and robustness of the FCN. The U-net architecture-based FCN framework was able to predict the type of nucleus class either belonging to the normal or abnormal classes from the cervical cancer smear images. It worked by assigning pixel-wise labels to individual nuclei in a whole slide image, which enabled the identification of multiple nuclei belonging to the same or different class as individual distinct instances.

[Fig. 5 about here.]

Wang et al. [137] conducted the nuclei segmentation process on cervical cancer morphological cell image using Mean-Shift clustering algorithm. Figure 6 shows the nuclei image processing and machine learning classification architecture by Wang et al. [137]. The classification was carried out based on the shape and textural features of the segmented images. The color space and Gabor features were extracted from the segmented image and put together to obtain a better classification performance. The nuclei segmentation-based analysis was also conducted by Rawat et. al [138] on leukaemia morphological cell image based on the global thresholding and histogram equalization. The details of nuclei image processing and machine learning classification architecture are described in Figure 7. The normal and abnormal classes were classified using the support vector machine (SVM) classifier. On the other hand, Negm et. al [139] conducted the K-Means clustering-based segmentation process on nucleus, cytoplasm and whole-cell of leukaemia morphological cell image to classify the normal and abnormal classes based on the decision support system classification.

[Fig. 6 about here.]

[Fig. 7 about here.]

[Fig. 8 about here.]

Figure 8 shows the architecture of nuclei image processing and machine learning classification by Negm et al. [139]. As described in this figure, K-means clustering segmentation process started by segmenting the nuclei or whole cell and the images were classified by a representation of three-color components, RGB (red, green, blue). The histogram of the color components indicates the contrast of the images. The most contrasted images were selected for K-mean clustering segmentation step. The K-means clustering-based segmentation was performed by partitioning the pre-processed image into K-mean clusters, classifying and grouping items into k groups (k is the number of pre-selected groups), minimizing the sum of squared distances between the items and the corresponding centroid used in grouping [140]. For example, if the grouping items are; background, other non-target cells and the cells to be extracted, thus, the K number is 3 (K1: background, K2: other non-targeted cells and K3: cells to be extracted).

Through the K-Means algorithm, the desired region of cells (nucleus, cytoplasm and whole-cell) can be separated from the unwanted region (background and other non-targeted cells). The segmented desired region can then proceed for features extraction step, which based on geometry, statistics, textures and size ratio. The analysis of these features was then performed to differentiate the regions for the classification step. Taken together, the techniques and algorithms used in cancer research are recommended to be used in stem cells research, particularly for biosafety and bioefficacy evaluation as summarized in Figure 9. Taking cues from the summary of the image processing pipeline in Figure 9, we proposed a framework-specific for biosafety and bioefficacy assessment, as depicted in Figure 10.

Ideally, this framework will be applicable to identify stem cells abnormality, particularly during the cell expansion phase. Following the proven studies in the similar domain, supervised learning will be employed where images with known normality level will be used as a training data. By utilizing CNN based image processing algorithm, the image of stem cells from microscope was segmented into smaller sub-image of a single cell that contains nuclei, cytoplasm and other features. In each segmented sub-image, the CNN based interpretation on the type of stem cell normality can be carried out according to the cell and nuclei features as exemplified previously [134]. The recommended features that should be considered are size, shape, grey level, color, texture, Law's Texture Energy (LTE), wavelet and Tamura's features. The performance metric for the model will be based on the features' percentage of stem cell from each single cell sub-image that showed normal or abnormal conditions. The sum of average for each feature will then become a metric for the classification using various models, such as Naive Bayes, Decision Tree or Random Forest. The summary of current machine learning application in stem cell research and cancer cell research is shown in Table 1.

[Table 1 about here.]

[Fig. 9 about here.]

[Fig. 10 about here.]

VI. THE CHALLENGES AND FUTURE OF PERSPECTIVES

In summary, the possibilities to adopt the aforementioned technical steps in stem cell research, particularly for risk evaluation in biosafety and bioefficacy are immense. The overlapping aspects of stem cell biology and cancer cell biology have led to the increase of large and highly complex datasets being generated from biological experiments from quantifying molecular variables, such as gene, protein and metabolites associated with different cancer and stem cells types. This has given insights into further understanding of the biological systems. Taken together, their involvement in disease progression and mechanism can be realized using machine learning and deep learning approaches. These approaches are able to address the complexity and heterogeneity of these datasets, providing new perspectives and generate novel hypotheses, particularly with regards to biosafety and bioefficacy risks and concerns in stem cells therapy. However, just as in any biomedical datasets, some of the challenges identified that may occur in stem cells research datasets are; 1) data requirements, which require large, labeled data to make deep learning successful, 2) overfitting in data training may inaccurately reflect underlying relationships, particularly in the heterogenous dataset and 3) interpretability of deep learning models may require better interpreting methods of its output [97].

Although the size of these datasets is increasing, there is still a need for massive, large datasets to reach meaningful perspectives and outcomes. Just as any biological system, data from stem cells biology can be incredibly complex with thousands of variables from different facets of physiological conditions. With suitable machine learning and deep learning models, we can assess the aspects of biosafety and bioefficacy of stem cells for clinical application. The generated model could also be used to identify fundamental design principles to create a suitable microenvironment for stem cells growth without jeopardizing their mortality and without altering their epigenetic components that may lead to cellular abnormality. However, to create such large and well-annotated datasets to study such complex network would require multi-omics datasets, which can be very expensive.

One of the options that could be utilized to take on this challenge is to use imaging data and analysis to characterize morphological and phenotypic changes of stem cells. This could be carried out by comparing the data from cancer and stem cells in various conditions and environmental perturbations as well as coupling it with deep learning algorithms. The data obtained would present interesting input in addressing biosafety and bioefficacy risks in stem cells therapy.

Nevertheless, we still have a long way to go to uncover and harnessing the potential of stem cells for therapy and to play a bigger role in the clinical settings. Machine learning approaches

offer another attractive alternative in understanding stem cells biology as well as their biosafety and bioefficacy risks. With the emerging developments of machine learning application, there will be new interpretation in the model features (as proposed in the aforementioned sections in this review). The interpretation could give rise to new and meaningful inputs or predictions from stem cell biological perspectives, which will be a challenge that is worth to explore.

ACKNOWLEDGMENTS

We would like to thank University of Malaya for providing the fund under the Research University (RU) Grant - Faculty Program (Project No: GPF012A-2018 and GPF039A-2019) and by the Ministry of Higher Education Malaysia (FRGS-FP114-2020). We thank reviewers and associate editor for their comments which improved this manuscript.

REFERENCES

- [1] M. Muraca, G. Galbiati, G. Realdi, M. Vilei, A. Fabricio, and M. Caruso, "Regenerative Medicine: An Insight," *Transplant Proc.*, vol. 39, no. 6, pp. 1995-1998, 2007.
- [2] A. S. Alvarado and S. Yamanaka, "Rethinking differentiation: stem cells, regeneration, and plasticity," *Cell*, vol. 157, no. 1, pp. 110-119, 2014.
- [3] B. A. Bunnell, M. Flaata, C. Gagliardi, B. Patel, and C. Ripoll, "Adipose-derived stem cells: isolation, expansion and differentiation," *Methods*, vol. 45, no. 2, pp. 115-120, 2008.
- [4] K. W. Yong et al., "Assessment of tumorigenic potential in long-term cryopreserved human adipose-derived stem cells," *J Tissue Eng Regen Med.*, vol. 11, no. 8, pp. 2217-2226, 2017.
- [5] N. Kalinina et al., "Characterization of secretomes provides evidence for adipose-derived mesenchymal stromal cells subtypes," *Stem Cell Res Ther.*, vol. 6, no. 221, pp. 1-12, 2015.
- [6] Y. Khoudj, J. Dawson, J. Coad, and L. Vona-Davis, "Hyperspectral imaging and K-means classification for histologic evaluation of ductal carcinoma in situ," *Front Oncol.*, vol. 8, no. 17, pp. 1-8, 2018.
- [7] L. Dawson et al., "Safety issues in cell-based intervention trials," *Fertil Steril*, vol. 80, no. 5, pp. 1077-1085, 2003.
- [8] F. J. Vizoso, N. Eiro, S. Cid, J. Schneider, and R. Perez-Fernandez, "Mesenchymal Stem Cell Secretome: Toward Cell-Free Therapeutic Strategies in Regenerative Medicine," *Int J Mol Sci.*, vol. 18, no. 9, 2017.
- [9] E. Mariani and A. Facchini, "Clinical applications and biosafety of human adult mesenchymal stem cells," *Curr Pharm Des.*, vol. 18, no. 13, pp. 1821-1845, 2012.
- [10] K. W. Yong, J. R. Choi, A. S. Dolbashid, and W. K. Wan Safwani, "Biosafety and bioefficacy assessment of human mesenchymal stem cells: what do we know so far?" *Regen Med.*, vol. 13, no. 2, pp. 219-232, 2018.
- [11] H. Yasrebi, P. Sperisen, V. Praz, and P. Bucher, "Can survival prediction be improved by merging gene expression data sets?" *PLoS One*, vol. 4, no. 10, pp. 11-13, 2009.
- [12] C. S. Huertas, L. G. Carrascosa, S. Bonnal, J. Valcárcel, and L. M. Lechuga, "Quantitative evaluation of alternatively spliced mRNA isoforms by label-free real-time plasmonic sensing," *Biosens Bioelectron.*, vol. 78, pp. 118-125, 2016.
- [13] T. Ching et al., "Opportunities and Obstacles for Deep Learning in Biology and Medicine," *J R Soc Interface*, vol. 15, no. 141, 2018.
- [14] H. Lian et al., "Integrative analysis of gene expression and DNA methylation through one-class logistic regression machine learning identifies stemness features in medulloblastoma," *Mol Oncol.*, vol. 13, no. 10, pp. 2227-2245, 2019.
- [15] M. Sherafatian and F. Arjmand, "Decision tree-based classifiers for lung cancer diagnosis and subtyping using TCGA miRNA expression data," *Oncol Lett.*, vol. 18, no. 2, pp. 2125-2131, 2019.
- [16] M. Quarta et al., "An artificial niche preserves the quiescence of muscle stem cells and enhances their therapeutic efficacy," *Nat Biotechnol.*, vol. 34, no. 7, pp. 752-759, 2016.
- [17] W. H. Land et al., "Kernelized partial least squares for feature reduction and classification of gene microarray data," *BMC Systems Biology*, vol. 5, no. 3, 2011.
- [18] D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, and J. J. Collins, "Next-Generation Machine Learning for Biological Networks," *Cell*, vol. 173, no. 7, pp. 1581-1592, 2018.
- [19] H. B. Sieburg, B. D. Rezner, and C. E. Muller-Sieburg, "Predicting clonal self-renewal and extinction of hematopoietic stem cells," *Proc Natl Acad Sci U S A*, vol. 108, no. 11, pp. 4370-4375, 2011.
- [20] J. Ritari et al., "Genomic prediction of relapse in recipients of allogeneic haematopoietic stem cell transplantation," *Leukemia*, vol. 33, no. 1, pp. 240-248, 2019.
- [21] N. Angayarkanni, D. Kumar, and G. Arunachalam, "The application of image processing techniques for detection and classification of cancerous tissue in digital mammograms," *J Pharm Sci Res.*, vol. 8, no. 10, pp. 1179-1183, 2016.
- [22] A. Ciurte, C. Selicean, O. Soritau, and R. Buiga, "Automatic detection of circulating tumor cells in darkfield microscopic images of unstained blood using boosting techniques," *PLoS One*, vol. 13, no. 12, pp. 1-11, 2018.
- [23] Y. Hamad, K. Simonov, and M. B. Naeem, "Breast cancer detection and classification using artificial neural networks," presented at the Proc - 2018 1st Annu Int Conf Inf Sci AiCIS 2018, Fallujah, Iraq, 2018.
- [24] M. M. Meenakshi, "Melanoma Skin Cancer Detection using Image Processing and Machine Learning," *Int J Trend Sci Res Dev.*, vol. 3, no. 4, pp. 780-784, 2019.
- [25] M. Zhang and J. M. Rosen, "Stem cells in the etiology and treatment of cancer," *Curr Opin Genet Dev.*, vol. 16, no. 1, pp. 60-64, 2006.
- [26] H. Niwa, T. Burdon, I. Chambers, and A. Smith, "Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3," *Genes & Development*, vol. 12, no. 13, pp. 2048-2060, 1998.
- [27] T. Burdon, I. Chambers, C. Stracey, H. Niwa, and A. Smith, "Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells," *Cells Tissues Organs*, vol. 165, no. 3-4, pp. 131-143, 1999.
- [28] S. Bao et al., "Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells," *Nature*, vol. 461, no. 7268, pp. 1292-1295, 2009.
- [29] M. Körbling and Z. Estrov, "Adult stem cells for tissue repair—a new therapeutic concept?" *N Engl J Med.*, vol. 349, no. 6, pp. 570-582, 2003.
- [30] D. T. Covas et al., "Multipotent mesenchymal stromal cells obtained from diverse human tissues share functional properties and gene-expression profile with CD146+ perivascular cells and fibroblasts," *Exp Hematol.*, vol. 36, no. 5, pp. 642-654, 2008.
- [31] S. W. Kim et al., "Successful Stem Cell Therapy Using Umbilical Cord Blood-Derived Multipotent Stem Cells for Buerger's Disease and Ischemic Limb Disease Animal Model," *Stem Cells*, vol. 24, no. 6, pp. 1620-1626, 2006.
- [32] F. Dazzi and N. J. Horwood, "Potential of mesenchymal stem cell therapy," *Curr Opin Oncol.*, vol. 19, no. 6, pp. 650-655, 2007.
- [33] M. T. Harting et al., "Intravenous mesenchymal stem cell therapy for traumatic brain injury," *J Neurosurg.*, vol. 110, no. 6, pp. 1189-1197, 2009.
- [34] F. Davatchi, B. S. Abdollahi, M. Mohyeddin, F. Shahram, and B. Nikbin, "Mesenchymal stem cell therapy for knee osteoarthritis. Preliminary report of four patients," *Int J Rheum Dis.*, vol. 14, no. 2, pp. 211-215, 2011.
- [35] Y. G. Koh and Y. J. Choi, "Infrapatellar fat pad-derived mesenchymal stem cell therapy for knee osteoarthritis," *Knee*, vol. 19, no. 6, pp. 902-907, 2012.
- [36] M. Stadtfeld, K. Brennand, and K. Hochedlinger, "Reprogramming of pancreatic β cells into induced pluripotent stem cells," *Curr Biol.*, vol. 18, no. 12, pp. 890-894, 2008.
- [37] M. Kennedy et al., "T lymphocyte potential marks the emergence of definitive hematopoietic progenitors in human pluripotent stem cell differentiation cultures," *Cell Rep.*, vol. 2, no. 6, pp. 1722-1735, 2012.
- [38] E. G. Z. Centeno, H. Cimarosti, and A. Bithell, "2D versus 3D human induced pluripotent stem cell-derived cultures for neurodegenerative disease modelling," *Mol Neurodegener.*, vol. 13, no. 1, p. 27, 2018.
- [39] R. Herbrecht et al., "Caspofungin first-line therapy for invasive

- aspergillosis in allogeneic hematopoietic stem cell transplant patients: An European Organisation for Research and Treatment of Cancer study," *Bone Marrow Transplant*, vol. 45, no. 7, pp. 1227-1233, 2010.
- [40] M. Piñeiro-Ramil et al., "Usefulness of Mesenchymal Cell Lines for Bone and Cartilage Regeneration Research," *Int J Mol Sci.*, vol. 20, no. 24, p. 6286, 2019.
- [41] J. Aziz, G. Liao, Z. Adams, M. Rizk, R. Shorr, and D. S. Allan, "Systematic review of controlled clinical studies using umbilical cord blood for regenerative therapy: Identifying barriers to assessing efficacy," *Cytotherapy*, vol. 21, no. 11, pp. 1112-1121, 2019.
- [42] Y. Sato et al., "Tumorigenicity assessment of cell therapy products: The need for global consensus and points to consider," *Cytotherapy*, vol. 21, no. 11, pp. 1095-1111, 2019.
- [43] M. Dominici et al., "Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement," *Cytotherapy*, vol. 8, no. 4, pp. 315-317, 2006.
- [44] Z. M. MacIsaac, H. Shang, H. Agrawal, N. Yang, A. Parker, and A. J. Katz, "Long-term in-vivo tumorigenic assessment of human culture-expanded adipose stromal/stem cells," *Exp Cell Res.*, vol. 318, no. 4, pp. 416-423, 2012.
- [45] K. W. Yong et al., "Phenotypic and functional characterization of long-term cryopreserved human adipose-derived stem cells," *Sci Rep.*, vol. 5, p. 9596, 2015.
- [46] W. K. Safwani, S. Makpol, S. Sathapan, and K. H. Chua, "Alteration of gene expression levels during osteogenic induction of human adipose derived stem cells in long-term culture," *Cell Tissue Bank*, vol. 14, no. 2, pp. 289-301, 2013.
- [47] W. K. Safwani, S. Makpol, S. Sathapan, and K. H. Chua, "Impact of adipogenic differentiation on stemness and osteogenic gene expression in extensive culture of human adipose-derived stem cells," *Arch Med Sci AMS.*, vol. 10, no. 3, pp. 597-606, 2014.
- [48] A. S. Lee, C. Tang, M. S. Rao, I. L. Weissman, and J. C. Wu, "Tumorigenicity as a clinical hurdle for pluripotent stem cell therapies," *Nat Med.*, vol. 19, no. 8, pp. 998-1004, 2013.
- [49] B. Blum and N. Benvenisty, "The tumorigenicity of human embryonic stem cells," *Adv Cancer Res.*, vol. 100, pp. 133-158, 2008.
- [50] W. K. Wan Safwani et al., "The effects of hypoxia and serum-free conditions on the stemness properties of human adipose-derived stem cells," *Cytotechnology*, vol. 68, no. 5, pp. 1859-1872, 2016.
- [51] J. R. Choi et al., "In situ normoxia enhances survival and proliferation rate of human adipose tissue-derived stromal cells without increasing the risk of tumorigenesis," *PLoS One*, vol. 10, no. 1, 2015.
- [52] J. R. Choi et al., "Impact of low oxygen tension on stemness, proliferation and differentiation potential of human adipose-derived stem cells," *Biochem Biophys Res Commun.*, vol. 448, no. 2, pp. 218-224, 2014.
- [53] K. H. Chua, F. Raduan, W. K. W. Safwani, N. F. Manzor, B. Pingguan-Murphy, and S. Sathapan, "Effects of serum reduction and VEGF supplementation on angiogenic potential of human adipose stromal cells in vitro," *Cell Prolif.*, vol. 46, no. 3, pp. 300-311, 2013.
- [54] C. K. Hui et al., "Human serum promotes the proliferation but not the stemness genes expression of human adipose-derived stem cells," *Biotechnology and Bioprocess Engineering*, vol. 17, no. 6, pp. 1306-1313, 2012.
- [55] X. Lei et al., "The effect of fluid shear stress on fibroblasts and stem cells on plane and groove topographies," *Cell Adh Migr.*, vol. 14, no. 1, pp. 12-23, 2020.
- [56] A. Hirano, M. Sano, N. Urushihata, H. Tanemura, K. Oki, and E. Suzuki, "Assessment of safety and feasibility of human allogeneic adipose-derived mesenchymal stem cells in a pediatric patient," *Pediatr Res.*, vol. 84, no. 5, pp. 575-577, 2018.
- [57] E. Lennmyr et al., "Survival in adult acute lymphoblastic leukaemia (ALL): A report from the Swedish ALL Registry," *Eur J Haematol.*, vol. 103, no. 2, pp. 88-98, 2019.
- [58] R. F. Schlenk et al., "Prospective Evaluation of Allogeneic Hematopoietic Stem-Cell Transplantation from Matched Related and Matched Unrelated Donors in Younger Adults with High-Risk Acute Myeloid Leukemia: German- Austrian Trial AMLHD98A," *J Clin Oncol.*, vol. 28, no. 30, pp. 4642-4648, 2010.
- [59] J. J. Cornelissen et al., "Results of a HOVON / SAKK donor versus no-donor analysis of myeloablative HLA- identical sibling stem cell transplantation in first remission acute myeloid leukemia in young and middle-aged adults: benefits for whom?," *Blood*, vol. 109, no. 9, pp. 3658-3666, 2007.
- [60] I. Kang et al., "Donor-dependent variation of human umbilical cord blood mesenchymal stem cells in response to hypoxic preconditioning and amelioration of limb ischemia," *Exp Mol Med.*, vol. 50, no. 4, p. 35, 2018.
- [61] K. H. Chua, W. K. Z. W. Safwani, A. A. Hamid, S. K. Shuhup, N. H. M. Haflah, and N. H. M. Yahaya, "Retropatellar fat pad-derived stem cells from older osteoarthritic patients have lesser differentiation capacity and expression of stemness genes," *Cytotherapy*, vol. 16, no. 5, pp. 599-611, 2014.
- [62] C. A. Herberts, M. S. Kwa, and H. P. Hermsen, "Risk factors in the development of stem cell therapy," *J Transl Med.*, vol. 9, no. 1, p. 29, 2011.
- [63] G. E. Georges, K. Doney, and R. Storb, "Severe aplastic anemia: allogeneic bone marrow transplantation as first-line treatment," *Blood Adv.*, vol. 2, no. 15, pp. 2020-2028, 2018.
- [64] J. M. Byun, D. H. Kim, D. Y. Shin, I. Kim, Y. Koh, and S. S. Yoon, "Combination of genetic aberration with international staging system classification for stratification of Asian multiple myeloma patients undergoing autologous stem cell transplantation," *In Vivo*, vol. 33, no. 2, pp. 611-619, 2019.
- [65] D. S. Jevotovsky, A. R. Alfonso, T. A. Einhorn, and E. S. Chiu, "Osteoarthritis and stem cell therapy in humans: a systematic review," *Osteoarthritis Cartilage*, vol. 26, no. 6, pp. 711-729, 2018.
- [66] C. Goswami, S. Poonia, L. Kumar, and D. Sengupta, "Staging system to predict the risk of relapse in multiple myeloma patients undergoing autologous stem cell transplantation," *Front Oncol.*, vol. 9, no. 633, pp. 1-9, 2019.
- [67] T. J. Myers, F. Granero-Molto, L. Longobardi, T. Li, Y. Yan, and A. Spagnoli, "Mesenchymal stem cells: Applications in cell and gene therapy," *Expert Opin Biol Ther.*, vol. 8, no. 3, pp. 97-122, 2009.
- [68] J. C. Ra et al., "Stem cell treatment for patients with autoimmune disease by systemic infusion of culture-expanded autologous adipose tissue derived mesenchymal stem cells," *J Transl Med.*, vol. 9, no. 181, pp. 1-11, 2011.
- [69] H. Fernández-Susavila, A. Bugallo-Casal, J. Castillo, and F. Campos, "Adult stem cells and induced pluripotent stem cells for stroke treatment," *Front Neurol.*, vol. 10, no. 908, 2019.
- [70] O. Y. Bang, E. H. Kim, J. M. Cha, and G. J. Moon, "Adult stem cell therapy for stroke: Challenges and progress," *J Stroke*, vol. 18, no. 3, pp. 256-266, 2016.
- [71] S. Shenoy et al., "Current Results and Future Research Priorities in Late Effects after Hematopoietic Stem Cell Transplantation for Children with Sickle Cell Disease and Thalassemia: A Consensus Statement from the Second Pediatric Blood and Marrow Transplant Consortium International Conference on Late Effects after Pediatric Hematopoietic Stem Cell Transplantation," *Biol Blood Marrow Transplant*, vol. 23, no. 4, pp. 552-561, 2017.
- [72] E. Meijer and J. J. Cornelissen, "Allogeneic stem cell transplantation in acute myeloid leukemia in first or subsequent remission: weighing prognostic markers predicting relapse and risk factors for non-relapse mortality," *Semin Oncol.*, vol. 35, no. 4, pp. 449-457, 2008.
- [73] A. Grigg and T. Hughes, "Role of Allogeneic Stem Cell Transplantation for Adult Chronic Myeloid Leukemia in the Imatinib Era," vol. 807, pp. 795-807, 2006.
- [74] T. Lodewyck and J. J. Cornelissen, "Allogeneic stem cell transplantation in acute myeloid leukemia: a risk-adapted approach," *Blood Rev.*, vol. 22, no. 6, pp. 293-302, 2008.
- [75] U. M. Saarinen-Pihkala et al., "Pathways Through Relapses and Deaths of Children with Acute Lymphoblastic Leukemia: Role of Allogeneic Stem-Cell Transplantation in Nordic Data," *Journal of Clinical Oncology*, vol. 24, no. 36, 2006.
- [76] A. Rovó and A. Tichelli, "Cardiovascular Complications in Long-Term Survivors After Allogeneic Hematopoietic Stem Cell Transplantation," *Semin Hematol.*, vol. 49, no. 1, pp. 25-34, 2012.
- [77] V. Volarevic et al., "Ethical and safety issues of stem cell-based therapy," *Int J Med Sci.*, vol. 15, no. 1, pp. 36-45, 2018.
- [78] D. C. Ding, Y. H. Chang, W. C. Shyu, and S. Z. Lin, "Review Human Umbilical Cord Mesenchymal Stem Cells: A New Era for Stem " Cell Therapy, vol. 24, no. 3, pp. 339-347, 2015.

- [79] B. J. Huntly et al., "MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors," *Cancer Cell*, vol. 6, no. 6, pp. 587-596, 2004.
- [80] N. S. Majhail et al., "Secondary solid cancers after allogeneic hematopoietic cell transplantation using busulfan- cyclophosphamide conditioning," *Blood*, vol. 117, no. 1, pp. 316-322, 2011.
- [81] C. R. Hake, T. A. Graubert, and T. S. Fenske, "Does autologous transplantation directly increase the risk of secondary leukemia in lymphoma patients?" *Bone Marrow Transplant*, vol. 39, no. 2, pp. 59-70, 2007.
- [82] J. Y. Lim et al., "Third-party regulatory T cells prevent murine acute graft-versus-host disease," *Korean J Intern Med.*, vol. 33, no. 5, pp. 980-989, 2018.
- [83] D. A. Van Valen et al., "Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments," *PLoS Comput Biol.*, vol. 12, no. 11, pp. 1-24, 2016.
- [84] Y. Atsuta et al., "Continuing increased risk of oral/esophageal cancer after allogeneic hematopoietic stem cell transplantation in adults in association with chronic graft- versus-host disease," *Ann Oncol.*, vol. 25, no. 2, pp. 435-441, 2014.
- [85] R. Or et al., "Safety and Efficacy of Donor Lymphocyte Infusions following Mismatched Stem Cell Transplantation," *Biol Blood Marrow Transplant*, vol. 12, no. 12, pp. 1295-1301, 2006.
- [86] S. M. Pollard et al., "Glioma Stem Cell Lines Expanded in Adherent Culture Have Tumor-Specific Phenotypes and Are Suitable for Chemical and Genetic Screens," *Cell Stem Cell*, vol. 4, no. 6, pp. 568-580, 2009.
- [87] Y. Lu et al., "Bone Marrow-Derived CD44+ Cells Migrate to Tissue-Engineered Constructs via SDF-1/CXCR4- JNK Pathway and Aid Bone Repair," *Stem Cells Int.*, vol. 2019, no. Article ID 1513526, p. 14, 2019.
- [88] E. Glaab, J. Bacardit, J. M. Garibaldi, and N. Krasnogor, "Using rule-based machine learning for candidate disease gene prioritization and sample classification of cancer gene expression data," *PLoS One*, vol. 7, no. 7, p. e39932, 2012.
- [89] A. R. G. Libby et al., "Automated Design of Pluripotent Stem Cell Self-Organization," *Cell Syst.*, vol. 9, no. 5, pp. 483-495, 2019.
- [90] M. A., "Machine learning predicts the look of stem cells," *Nature*.
- [91] H. Zhang et al., "A novel machine learning based approach for iPS progenitor cell identification," *PLoS Comput Biol.*, vol. 15, no. 12: e1007351, pp. 1-19, 2019.
- [92] P. K. Mallick, S. H. Ryu, S. K. Satapathy, S. Mishra, G. N. Nguyen, and P. Tiwari, "Brain MRI image classification for cancer detection using deep wavelet autoencoder-based deep neural network," *IEEE Access*, vol. 7, pp. 46278-46287, 2019.
- [93] N. Khuriwal and N. Mishra, "Breast Cancer Diagnosis Using Deep Learning Algorithm," presented at the 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida (UP), India, 2018.
- [94] J. Taninaga et al., "Prediction of future gastric cancer risk using a machine learning algorithm and comprehensive medical check-up data: A case-control study," *Scientific Reports*, vol. 9, no. 1, pp. 1-9, 2019.
- [95] S. Webb, "Deep learning for biology," *Nature*, vol. 554, no. 7693, pp. 555-557, 2018.
- [96] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med Image Anal.*, vol. 42, pp. 60-88, 2017.
- [97] A. A. Kalinin et al., "Deep learning in pharmacogenomics: From gene regulation to patient stratification," *Pharmacogenomics*, vol. 19, no. 7, pp. 629-650, 2018.
- [98] F. Ke, Z. Sheng, Z. Ying, L. Jun, M. Holcombe, and Z. Xiao, "A Machine Learning Assisted, Label-free, Non-invasive Approach for Somatic Reprogramming in Induced Pluripotent Stem Cell Colony Formation Detection and Prediction," *Scientific Reports*, vol. 7, no. 1, pp. 1-9, 2017.
- [99] N. Tanaka, T. Yamashita, A. Sato, V. Vogel, and Y. Tanaka, "Simple agarose micro-confinement array and machine-learning-based classification for analyzing the patterned differentiation of mesenchymal stem cells," *Plos One*, vol. 12, no. 4, p. e0173647, April 5, 2017 2017.
- [100] R. Theagarajan and B. Bhanu, "DeepHesc 2.0: Deep generative multi adversarial networks for improving the classification of hESC," *PLoS One.*, vol. 14, no. 3, pp. 1-28, 2019.
- [101] M. Shackleton, "Normal stem cells and cancer stem cells: Similar and different," *Semin Cancer Biol.*, vol. 20, no. 2, pp. 85-92, 2010.
- [102] C. J. Bradhurst, W. Boles, and Y. Xiao, "Segmentation of bone marrow stromal cells in phase contrast microscopy images," presented at the 2008 23rd International Conference Image and Vision Computing New Zealand, Christchurch, 2008.
- [103] C. Jiang and K. Tsai, "Image segmentation techniques for stem cell tracking," presented at the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, 2013.
- [104] V. Tirino et al., "Human primary bone sarcomas contain CD133 + cancer stem cells displaying high tumorigenicity in vivo," *FASEB J.*, vol. 25, no. 6, pp. 2022-2030, 2011.
- [105] P. B. Gupta, C. L. Chaffer, and R. A. Weinberg, "Cancer stem cells: Mirage or reality?," *Nat Med.*, vol. 15, no. 9, pp. 1010-1012, 2009.
- [106] C. M. Megyola et al., "Dynamic migration and cell-cell interactions of early reprogramming revealed by high- resolution time-lapse imaging," *Stem Cells*, vol. 31, no. 5, pp. 895-905, 2013.
- [107] O. Dzyubachyk et al., "Automated analysis of time- lapse fluorescence microscopy images: from live cell images to intracellular foci," *Bioinformatics*, vol. 26, no. 19, pp. 2424- 2430, 2010.
- [108] E. Turetken, X. Wang, C. J. Becker, C. Haubold, and P. Fua, "Network Flow Integer Programming to Track Elliptical Cells in Time-Lapse Sequences," *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 942-951, 2017.
- [109] F. Papaccio, F. Paino, T. Regad, G. Papaccio, V. Desiderio, and V. Tirino, "Concise Review: Cancer Cells, Cancer Stem Cells, and Mesenchymal Stem Cells: Influence in Cancer Development," *Stem Cells Transl Med.*, vol. 6, no. 12, pp. 2115-2125, 2017.
- [110] A. V. Suhovskih, V. I. Kashuba, G. Klein, and E. V. Grigorieva, "Prostate cancer cells specifically reorganize epithelial cell-fibroblast communication through proteoglycan and junction pathways," *Cell Adhes Migr.*, vol. 11, no. 1, pp. 39-53, 2017.
- [111] T. J. Liu et al., "CD133 + cells with cancer stem cell characteristics associates with vasculogenic mimicry in triple- negative breast cancer," *Oncogene*, vol. 32, no. 5, pp. 544-553, 2013.
- [112] A. Biddle et al., "Cancer stem cells in squamous cell carcinoma switch between two distinct phenotypes that are preferentially migratory or proliferative," *Cancer Res.*, vol. 71, no. 15, pp. 5317-5326, 2011.
- [113] S. Y. Koh, J. Y. Moon, T. Unno, and S. K. Cho, "Baicalein suppresses stem cell-like characteristics in radio- and chemoresistant MDA-MB-231 human breast cancer cells through up-regulation of IFIT2," *Nutrients*, vol. 11, no. 3:624, 2019.
- [114] F. Brugnoli, S. Grassilli, Y. Al-Qassab, S. Capitani, and V. Bertagnolo, "CD133 in Breast Cancer Cells: More than a Stem Cell Marker," *J Oncol.*, pp. 1-8, 2019.
- [115] Y. Xu, C. So, H. M. Lam, M. C. Fung, and S. Y. Tsang, "Apoptosis Reversal Promotes Cancer Stem Cell-Like Cell Formation," *Neoplasia* vol. 20, no. 3, pp. 295-303, 2018.
- [116] A. Herreros-Pomares et al., "Lung tumorspheres reveal cancer stem cell-like properties and a score with prognostic impact in resected non-small-cell lung cancer," *Cell Death Dis.*, vol. 10, no. 9:660, pp. 1-14, 2019.
- [117] M. Singla, A. Kumar, A. Bal, S. Sarkar, and S. Bhattacharyya, "Epithelial to mesenchymal transition induces stem cell like phenotype in renal cell carcinoma cells," *Cancer Cell Int.*, vol. 18, no. 57, pp. 1-13, 2018.
- [118] Q. Jiang et al., "A fraction of CD133+ CNE2 cells is made of giant cancer cells with morphological evidence of asymmetric mitosis," *J Cancer*, vol. 6, no. 12, pp. 1236-1244, 2015.
- [119] W. Lim et al., "Association between cancer stem cell-like properties and epithelial-to-mesenchymal transition in primary and secondary cancer cells," *Int J Oncol.*, vol. 49, no. 3, pp. 991-1000, 2016.
- [120] K. M. Joo, S. Y. Kim, and X. Jin, "Clinical and biological implications of CD133-positive and CD133- negative cells in glioblastomas," *Lab Invest.*, vol. 88, no. 8, p. 808, 2008.
- [121] A. Eramo et al., "Identification and expansion of the tumorigenic lung cancer stem cell population," *Cell Death Differ.*, vol. 15, no. 3, pp. 504-514, 2008.
- [122] P. Gener et al., "Pivotal role of AKT2 during dynamic phenotypic change of breast cancer stem cells," *Cancers (Basel)*, vol. 11, no. 8, pp. 1-18, 2019.
- [123] D. Kusumoto et al., "Automated Deep Learning- Based System to Identify Endothelial Cells Derived from Induced Pluripotent Stem

- Cells," Stem Cell Reports, vol. 10, no. 6, pp. 1687-1695, 2018.
- [124] A. Waisman et al., "Deep Learning Neural Networks Highly Predict Very Early Onset of Pluripotent Stem Cell Differentiation," Stem Cell Reports, vol. 12, no. 4, pp. 845- 859, 2019.
- [125] F. Buggenthin et al., "Prospective identification of hematopoietic lineage choice by deep learning," Nature Methods, vol. 14, no. 4, pp. 403-406, 2017.
- [126] V. Adamski, A. Hempelmann, C. Flüh, R. Lucius, M. Synowitz, and K. Hattermann, "Dormant glioblastoma cells acquire stem cell characteristics and are differentially affected by Temozolomide and AT101 treatment," Oncotarget, vol. 8, no. 64, p. 108064, 2017.
- [127] N. Savage, "How AI is improving cancer diagnostics," Nature, vol. 579, pp. S14-S16, 2020.
- [128] E. Landhuis, "Deep learning takes on tumours," Nature, vol. 580, no. 7804, pp. 551-553, 2020.
- [129] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," Scientific Reports, vol. 9, p. 12495, 2019.
- [130] T. Saba, "Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges," Journal of Infection and Public Health, vol. 13, no. 9, pp. 1274-1289, 2020.
- [131] C. Scheeder, F. Heigwer, and M. Boutros, "Machine learning and image-based profiling in drug discovery," Current Opinion in Systems Biology, vol. 10, pp. 43-52, 2018.
- [132] Z. Zhang et al., "Similarity in gene-regulatory networks suggests that cancer cells share characteristics of embryonic neural cells," J Biol Chem., vol. 292, no. 31, pp. 12842-12859, 2017.
- [133] L. Y. e. al, "Deep learning," Nature, vol. 521, pp. 436-444, 2015.
- [134] S. S. Naik and A. Dixit, "Cancer detection using image processing and machine learning," International Journal of Engineering Research & Technology vol. 8, no. 6, pp. 400-402, 2019.
- [135] M. I. e. al., "Challenges in Clinical Development of Mesenchymal Stromal/Stem Cells: Concise Review," Stem Cells Translational Medicine, vol. 8, pp. 1135 – 1148, 2019.
- [136] E. Hussain, L. B. Mahanta, C. R. Das, M. Choudhury, and M. Chowdhury, "A shape context fully convolutional neural network for segmentation and classification of cervical nuclei in Pap smear images," Artificial Intelligence in Medicine, vol. 107, p. 101897, 2020.
- [137] P. Wang, L. Wang, Y. Li, Q. Song, S. Lv, and X. Hu, "Automatic cell nuclei segmentation and classification of cervical Pap smear images," Biomedical Signal Processing and Control, vol. 48, pp. 93-103, 2019.
- [138] J. Rawat, A. Singh, B. Hs, J. Virmani, and J. S. Devgun, "Computer assisted classification framework for prediction of acute lymphoblastic and acute myeloblastic leukemia," Biocybernetics and Biomedical Engineering vol. 37, no. 4, pp. 637-659, 2017.
- [139] A. Negm, O. A. Hassan, and A. Kandil, "A decision support system for Acute Leukaemia classification based on digital microscopic images," Alexandria Engineering Journal vol. 57, pp. 2319-2332, 2018.
- [140] X. Wu et al., "Top 10 algorithms in data mining," Knowledge and Information Systems, vol. 14, no. 1, pp. 1-37, 2008.



WAN SAFWANI WAN KAMARUL ZAMAN

received her Master of Pharmacy (MPharm) from the University of Strathclyde, United Kingdom in 2003. In 2006, she received her MSc in Immunopharmacology, also from the University of Strathclyde, United Kingdom. In 2008, she was awarded the National Science Fellowship (NSF) from the Ministry of Science,

Technology & Innovation (MOSTI), Malaysia allowing her to pursue and complete her PhD in 2012 at Universiti Kebangsaan Malaysia (National University of Malaysia). She currently serves as a senior lecturer in the Department of Biomedical Engineering, Faculty of Engineering, University of Malaya and as an executive member of the Tissue Engineering and Regenerative Medicine Society of Malaysia (TESMA). She has contributed substantially in terms of publication and as a reviewer for various journals. Her research interest is on the study of bio-safety and bio-efficacy of adult stem cells and stem cells niche regulation.



SALMAH BT. KARMAN (Member, IEEE) received her Bachelor's degree in Electrical and Electronic Engineering from Oita University, Japan in 2003. She then received Master in Engineering Science (M.Eng.Sc) from University of Malaya, Kuala Lumpur, Malaysia in 2011. In 2016, she received her PhD from Universiti Kebangsaan Malaysia (National University of Malaysia). Currently, she is working as a senior lecturer in the Department of Biomedical Engineering, Faculty of Engineering, University of Malaysia. Her research interest mostly focuses on biosensor and bioinspired devices.



MOHD YAZED B. AHMAD (Member, IEEE) received the B.E degree in Electrical Engineering from the Department of Electrical Engineering, University of Malaya, Kuala Lumpur, Malaysia in 2003. He then received M. Eng. Sc (Instrumentation) from the Department of Biomedical Engineering, University of Malaya in 2006 and PhD in Electrical

Engineering Science from the University of Technology Sydney, NSW, Australia in 2013. He is currently a senior lecturer and a programme coordinator at the Department of Biomedical Engineering, Faculty of Engineering, University of Malaya. His current research interests include instrumentation & real time machine learning system, wireless power transfer, smart space and IOT system.



EFFIRUL IKHWAN RAMLAN received his B. Sc. (Hons) degree in Intelligent Systems from Universiti Teknologi Mara (UiTM), Malaysia. He earned a PhD in Computer Science from University of Southampton, UK (in 2009) for computational and theoretical development of nucleic acid computers as information processors. He is currently a lecturer at the

School of Computing, Engineering and Intelligent Systems (SCEIS), Ulster University, UK. His research focuses on the development of synthetic devices for biological computing, machine learning approaches in genomics and unconventional bio-inspired algorithms for multi-objective optimisation problems.



SITI NURAINIE BT TUKIMIN received her Bachelor's degree (B. Sc.) in Science of Biology from Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia in 2014. She is currently pursuing her Master's degree in Engineering of Science from Department of Biomedical Engineering, Faculty of Engineering, University of Malaya (UM), Kuala Lumpur, Malaysia. She is currently doing research on tumor/cancer cell morphology and characteristics for development of tumor tissue mimicking phantom.

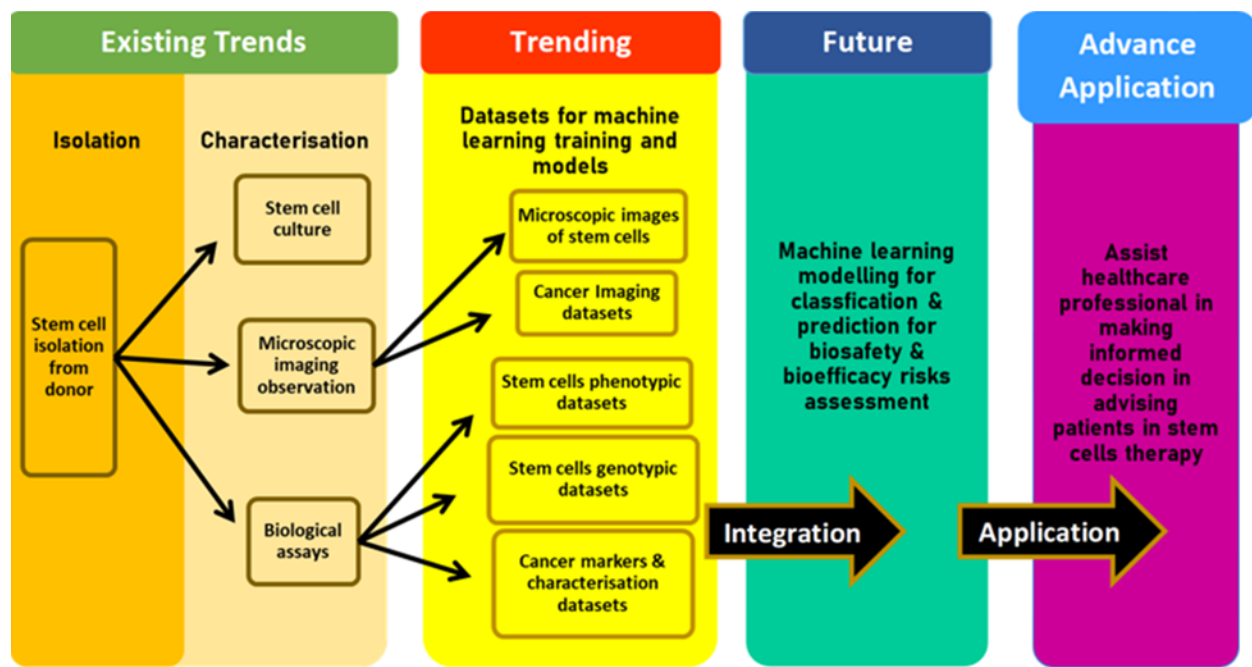


FIGURE 1: An overview of the current research trend in stem cells research. Collection of datasets particularly on characterization of stem and cancer cells would enable machine learning approach to be applied in stem cells research and subsequently in stem cells therapy.

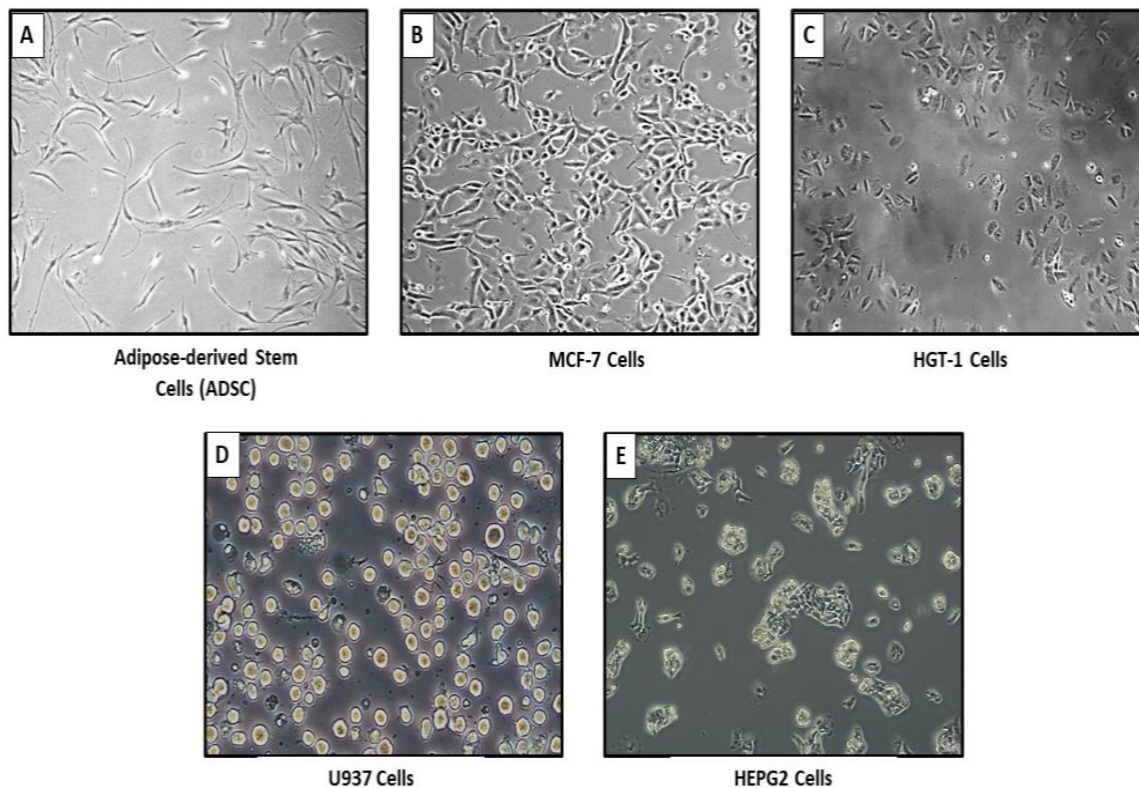


FIGURE 2: Representative images of human adult stem cells and selected cancer cell lines (10X magnification). A) ADSC (Human adipose-derived stem cells), B) MCF-7 (Human breast cancer cell line), C) HGT-1 (Human gastric cancer cell line), D) U937 (Human lymphoma cell line) and E) HEPG2 (Human liver cancer cell line). Differences in cell morphology may enable machine learning approach in evaluating the risks of biosafety and bioefficacy in stem cells therapy.

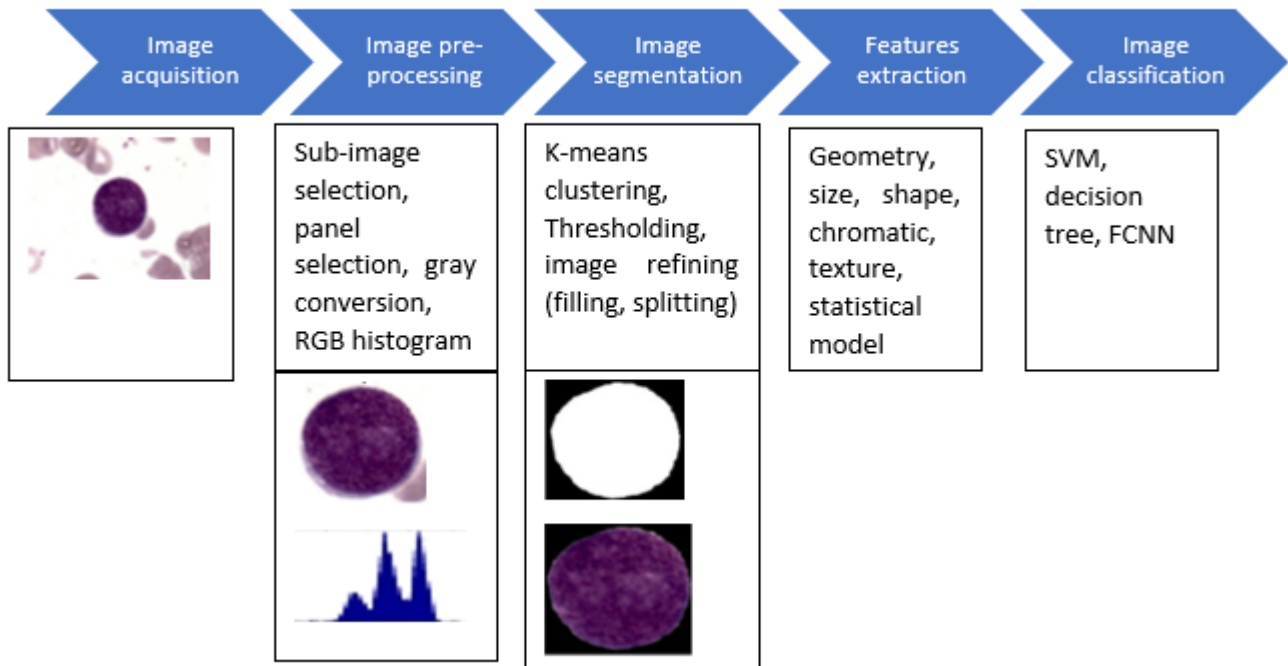


FIGURE 3: Overview of a typical pipeline of image processing steps for machine learning classification. Adapted from Jyoti Rawat et. al [138].

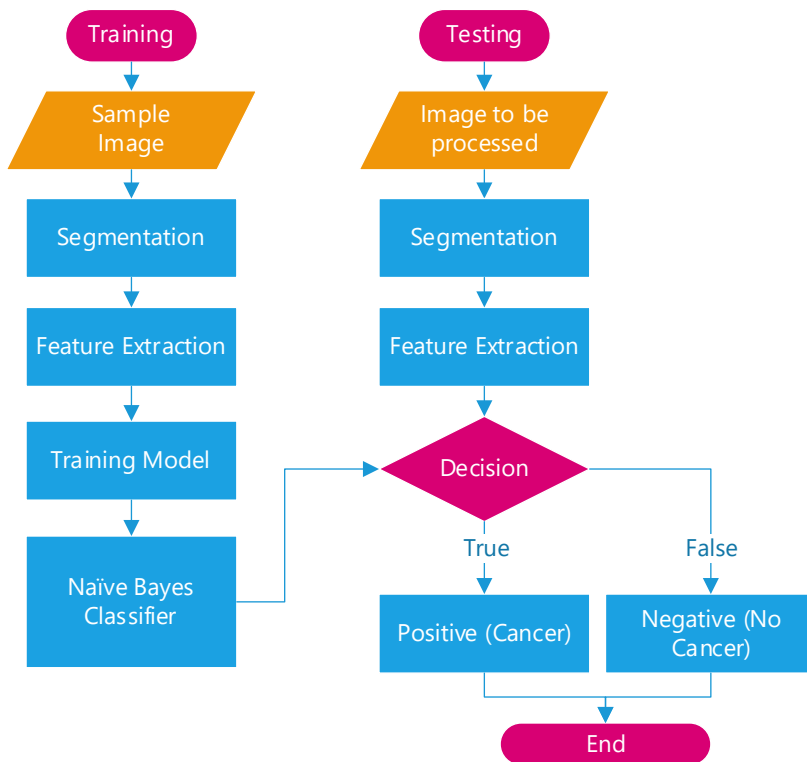


FIGURE 4: The machine learning architecture of cancer detection process which can be applied in stem cells biosafety and bioefficacy assessment. Adapted from Naik & Dixit [134].

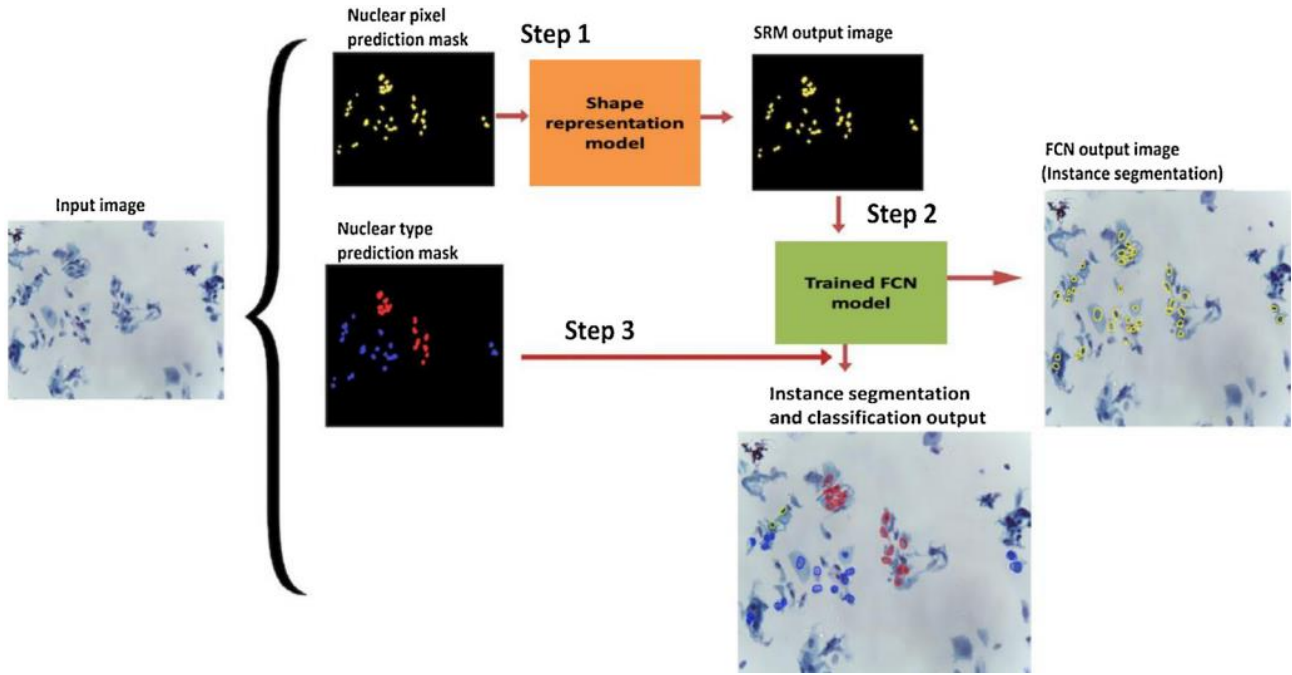


FIGURE 5: Nuclei image processing and machine learning classification architecture using U-net architecture-based fully convolutional neural network (FCN). Reproduced from Hussain et al. [136] with permission.

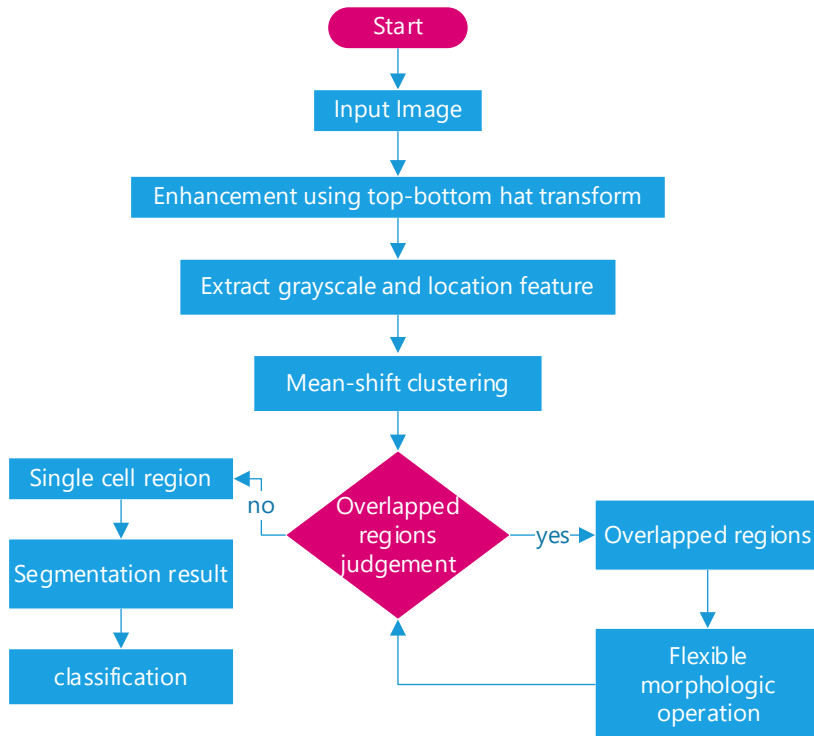


FIGURE 6: Nuclei image processing and machine learning classification architecture using mean-shift clustering algorithm. Reproduced from Wang et al. [137] with permission

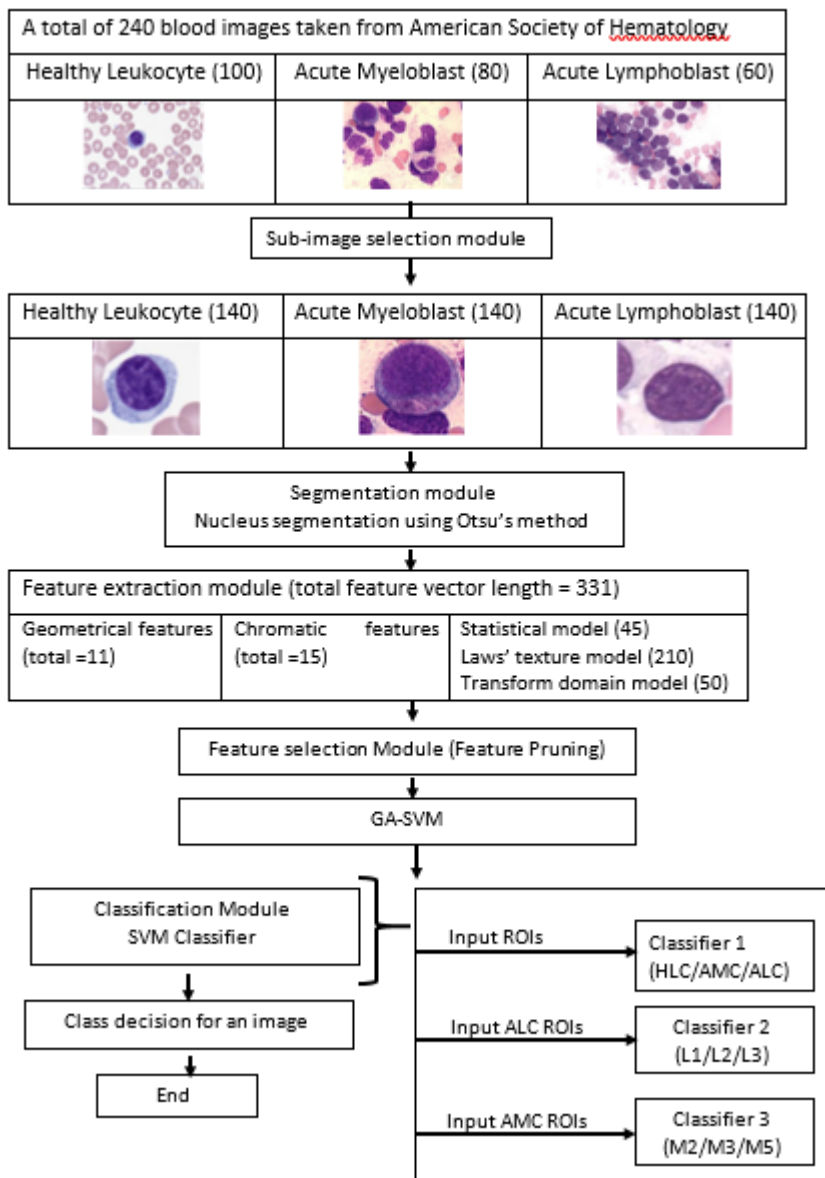


FIGURE 7: Nuclei image processing and machine learning classification architecture using support vector machine (SVM) classifier module. Reproduced from Jyoti Rawat et. al [138] with permission.

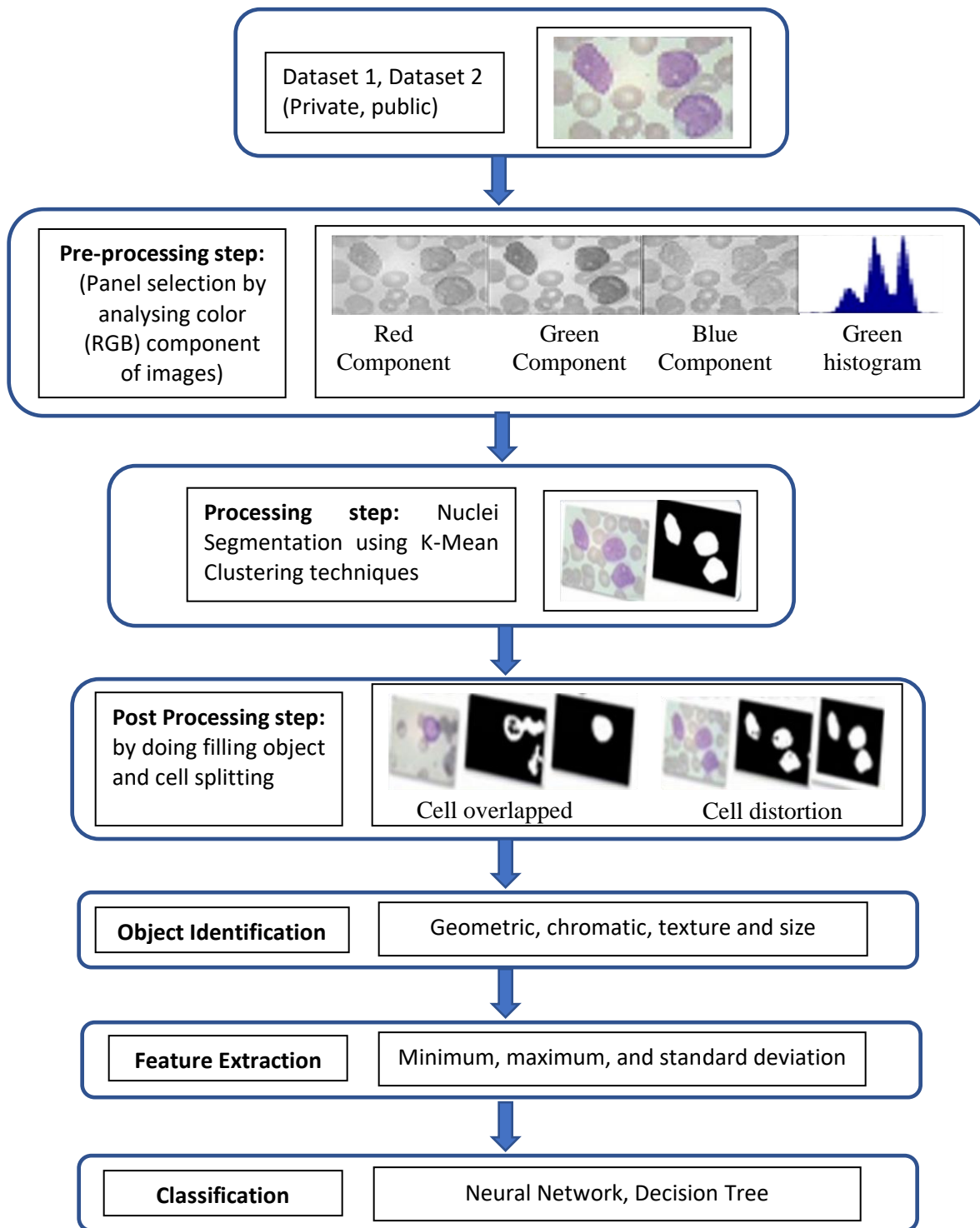


FIGURE 8: Nuclei image processing and machine learning classification architecture using K-Means clustering-based segmentation. Reproduced from Negm et al. [139] with permission.

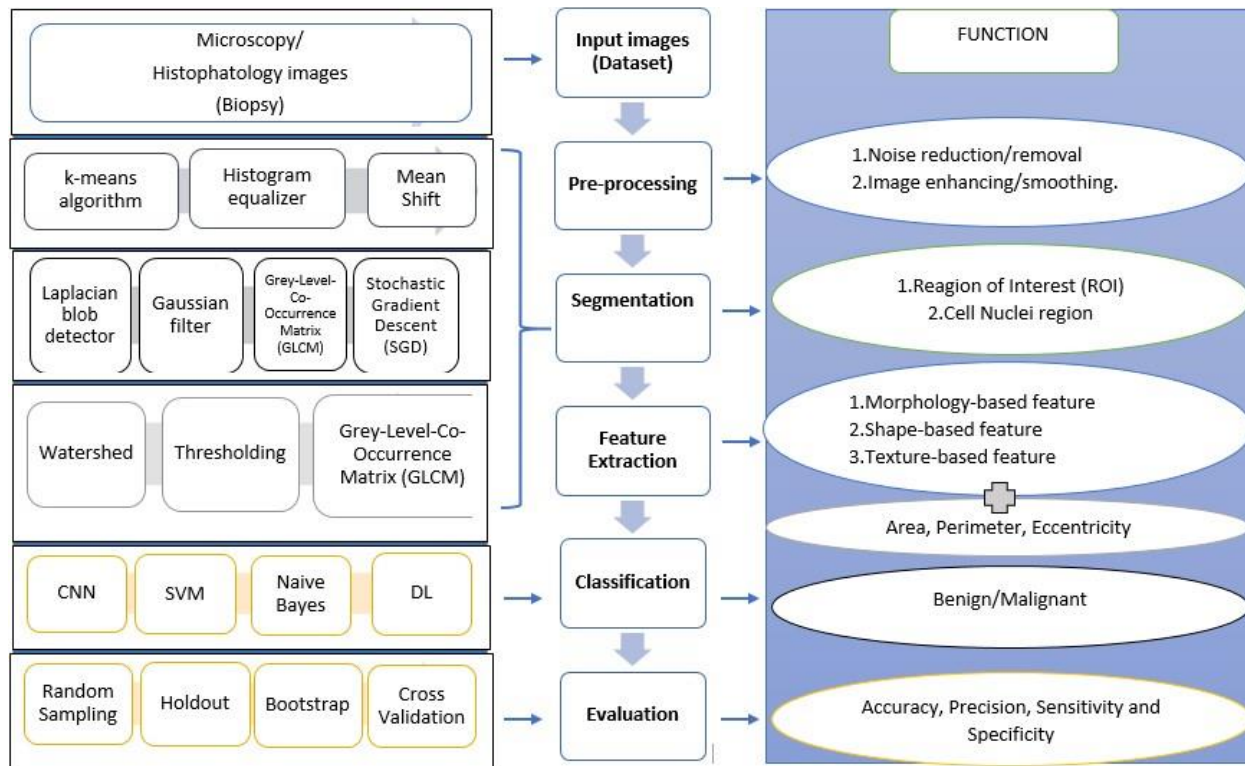


FIGURE 9: Overview of commonly used image processing-based machine learning techniques in cancer research which can be applied in stem cell research for stem cells biosafety and bioefficacy assessment.

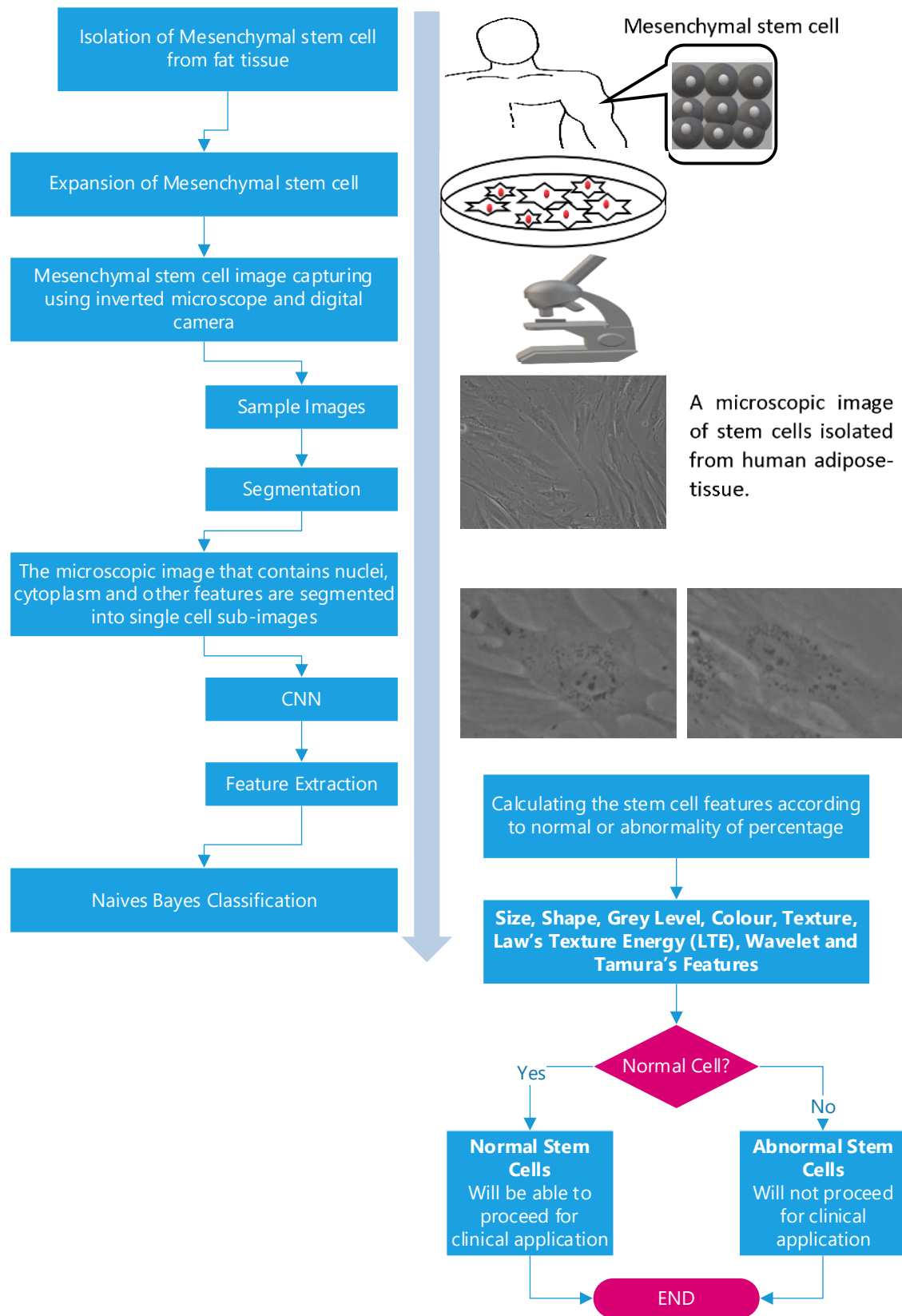


FIGURE 10: Image processing machine learning architecture proposed for stem cells research particularly for biosafety and bioefficacy assessment using microscopy images of human adipose derived stem cells as an example.

TABLE 1

A SUMMARY OF CURRENT MACHINE LEARNING APPLICATION IN STEM CELL RESEARCH AND CANCER RESEARCH.

Ref.	Type of Cell	Task	Learning Method	Feature Selection	Parameters
[91]	iPSC	Live-cell microscopic imaging dataset for iPS progenitor cells detection.	XGBoost algorithm	11 types of features are extracted (volume, area, sphericity, ellipsoid-prolate, ellipsoid-oblate, nucleus-cytoplasm volume ratio, displacement, speed, Intensity-StdDev, Intensity-Max, Intensity-Min) using time window and two-step feature selection.	XGBoost setting (learning rate = 0.01, n_estimators = 385, gamma = 0, and 5-Kfold cross validation).
[98]	iPSC	Time-lapse bright field image analysis for iPS colony formation detection and prediction	CNN for iPSC colonies recognition, HMM for growth curve modelling (Baum-Welch and Viterbi algorithms), and AlexNet algorithm as classifier.	2 main types of features selected (Colony texture and area of growth) using the sliding window method with Autolevels (AL) algorithm, Gaussian filter (GFP) and Random Walker algorithm for dataset pre-processing.	Batch normalized layer training for AlexNet with manual human validation.
[123]	iPSC	Phase contrast and immunofluorescence image analysis of differentiated endothelial cells	LeNet and AlexNet algorithm	2 types of feature selected (Differentiated feature of endothelial cells – vascular tube and CD31 expression)	LeNet and AlexNet architectures are available in [123] (Figure 3). Training was conducted with stochastic gradient descent, learning rate = 0.01, and cross-entropy error as loss function. Weight initialization using Xazier algorithm with dropout-techniques applied to avoid overfitting.
[100]	Human ESC	Phase contrast microscopic ESC images classification	DeephESC 2.0 algorithm consisting of CNN, Triplet CNNs and Fused CNN-Triplet.	6 types of features selected (cell clusters, debris, unattached cells, attached cells, dynamically blebbing cells and apoptically blebbing cells) following the image pre-processing with intensity modeling of a mixture of two Gaussians. Generative Multi Adversarial Networks (GMAN) is implemented for synthetic hESC images dataset.	DeephESC 2.0 setting (learning rate = 1.2×10^{-2} , momentum = 0.9 and weight decay = 1×10^{-3}), variation of architectures to improve hESC classification is available in [100] (Table 4). Two triplets CNN were introduced to perform fine-grained classification using the architecture presented in Figure 9 [100] with hyper-parameters setting in Table 6. GMAN discriminators setting is available in Table 7 [100].
[99]	MSC	Microphotograph of MSC image for differentiation analysis	SVM algorithm	3 main types of features selected (Adipogenic differentiation, osteogenic differentiation and non-differentiated) based on the intensity balance of RGB channels.	RGB pixel intensities with maximum value of R = 255, G = 255, and B = 255.
[107]	Mouse ESC	Time-lapse fluorescence microscopic image analysis of cell foci and cell motion	Level-set-based cell segmentation and tracking algorithm	1 type of feature selected (time course in cell phase transition) using cell segmentation and tracking with cell motion correction and foci segmentation and pattern recognition.	Level-set-based cell segmentation and tracking hyperparameters (motion compensation algorithm, $\alpha = 0.5$ and $\beta = 0.95$; λ of the smoothness energy term of the foci segmentation algorithm = 10%; focus size = 3-5 voxels; expected contrast = 0.05 & 0.2; cell-phase classification threshold = 30% of foci at boundary)
[129]	Breast cancer	Digital mammography images classification	CNN algorithm (fully connected (FC) layers involving VGG network or Resnet network as	8 types of feature selected (regional area, major axis length, mean intensity, background, malignant mass, benign mass, malignant calcification and benign	CNN setting (First step with 3 training strategy: learning rate = 10^{-3} , 10^{-4} , 10^{-5} . Second step with 2 training strategy: learning rate = to 10^{-4} ; weight

			patch classifier layers). The complete network design is presented in Figure 1 [129].	calcification).	decay = 0.001 and learning rate to 10^{-5} ; weight decay = 0.01)
[134]	Breast cancer	Magnetic Resonance Images (MRI) dataset classification	CNN algorithm for image processing and Naïve Bayes as classifier	4 types of colour-based feature selected (grey-level texture feature, Law's Texture energy, Tamura feature and wavelet feature)	N/A
[136]	Cervical cancer	Pap smear images classification	FCN algorithm with U-net variant comprising of residual blocks from Deep Residual Networks and dense blocks from Densely Connected Networks within the shape representation model (SRM) encoder.	3 types of feature selected (cell nuclei irregularity size, shape and texture). Stacked auto-encoder (SAE) is implemented to construct SRM for segmentation.	FCN network architecture is presented in Figure 4 [136]. Optimization of the network was performed with Adam optimizer. FCN algorithm hyperparameters (stochastic gradient descent, learning rate = 0.01; reduced by a factor of 10 every 5 epochs, and momentum = 0.88 for 200 epochs)
[137]	Cervical cancer	Pap smear images classification	SVM algorithm as classifier and Mean-shift clustering algorithm with mathematical morphology for segmentation.	5 types of selected features (cell nuclei area, perimeter, eccentricity, roundness, circularity) Optimal feature set is obtained using chain-like agent genetic algorithm (CAGA), P-value and maximum relevance-minimum multicollinearity (MRmMC).	N/A
[138]	Lymphoblastic and Myeloblastic leukemia	Microscopic blood images classification	Genetic algorithm (GA) for feature selection with SVM algorithm as classifier.	3 main types of selected features (colour, statistical texture & geometric - area, perimeter, diameter, Euler's no, major axis, minor axis, solidity, eccentricity, roundness, convex area and extent). The complete list of the extracted features and methods for extraction is available in Table 6 [138].	GA setting (size population = 20, number of generations = 10, replacement rate = 0.8, crossover fraction = 0.5, mutation fraction = 0.01, fitness scaling = rank, selection function = roulette, no of variables = 331), SVM hyperparameters (Gamma g = 0.0057/ κ , Cost C = 15,334). Various kernel functions and hyperparameters configuration were experimented. Details of the recommended configurations are available in Table 15, with details analysis in Table 8 - 14 [138].
[139]	Leukaemia	Acute leukemia blast cells in colored microscopic images classification	Artificial Neural Network (ANN) and Decision Tree as classifier with K-means clustering algorithm	4 main types of selected features (geometry, statistics, textures, and size ratio from selected regions in nucleus, cytoplasm and whole cell)	N/A

Notes; iPSC-induced Pluripotent Stem Cells, MSC-Mesenchymal Stem Cells, ESC-Embryonic Stem Cells, Min-Minimum, Max-Maximum